September, 2021

Research Software Current State Assessment

Authored by the Alliance's Advanced Research Software Working Group:

Qian Zhang, Susan Brown, James Colliander, Brian Corrie, Gabor Fichtinger, Scott Henwood, Mark Leggott, Catherine Lovekin, Felipe Pérez-Jvostov, Ghilaine Roquet, and Marc-Étienne Rousseau

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Table of Contents

1 Executive Summary

As the first position paper on research software (RS) in Canada, this report surveys and summarizes RS generally as an emerging field and as an area of professionalization, nationally and internationally, and documents strengths, challenges, and opportunities within the current RS ecosystem, as they pertain to the Digital Research Alliance of Canada (the Alliance). The purpose of this paper is to frame an understanding of the RS landscape and establish a general framework for conversation within Canada.

This report complements The Current State of Research Data Management (RDM) in Canada[1] and The Current State of Advanced Research Computing (ARC) in Canada[2] that update the 2017 DM and ARC position papers submitted to Innovation, Science and Economic Development Canada (ISED) by the Leadership Council for Digital Research Infrastructure (LCDRI). At that time RS was recognized as an essential part of the digital landscape; however, the area lacked the maturity and structured communities that enabled the production of the other two papers. This RS report therefore represents the first current state assessment of RS in Canada.

This report is intended to allow the Alliance to understand and build on the current state and facilitate a strategy that advances RS in coordination with other digital research infrastructure (DRI) elements to support research excellence in Canada. Findings and observations in this document, alongside the RDM and ARC Current State Assessment publications, are meant to provide background information to the Alliance analysts and management, the Alliance Board, and the Alliance Researcher Council, to support the development of the Alliance's New Service Delivery Model (NSDM), Strategic Planning, and Funding Model Delivery.

RS is multidimensional and international. For this document, RS is defined from a broad perspective at an emerging and complex intersection of tools, disciplines, services, platforms, hardware, resources, and the people (users, researchers, developers, stakeholders, personnel, communities, etc.) who use and contribute to them. RS is fundamental to research, given its role in helping researchers make new discoveries and share their outputs with others. Both technological and cultural trends in DRI are driving the continual and rapid evolution of RS, making it even more critical in advancing research. Transformation to an open science-driven[3] RS culture depends on the creation of tools, platforms and services that enable researchers to mobilize knowledge and make research processes more efficient, transparent, reproducible, and responsive to societal challenges. Specific elements of this shift include: increasing collaboration and interaction among researchers; the development of technical infrastructure that promotes the adoption of emerging research practices; the development, promotion and adoption of open-source and open science practices. All of those shifts require an agile and responsive ecosystem

---

[1] https://engagedri.ca/assets/documents/NDRIO_RDM_PositionPaper.pdf

[2] Unpublished document

[3] Open science in this document is a broad term to indicate a wide range of activities and forms of openness in the scholarly and research environment that apply to all disciplines, including open access, open data, open source, open reproducible research, open governance, open science evaluation, open peer review, open educational resource, open collaboration, open education, citizen science, etc.

with strong, highly qualified personnel (HQP) support and sustainable funding if Canada is to meet 21st-century challenges and remain competitive internationally.

RS is a DRI component that is tightly coupled to ARC and RDM, as recognized in the federal government's DRI strategy, regardless of whether one is analyzing RS characteristics, types, functions, needs, or impact on digital research. Within the DRI ecosystem the fundamental elements of RS, RDM, ARC, and cybersecurity are fundamentally interconnected and interdependent. Viewing the research enterprise through the lenses of the RS, data, and research lifecycles reveals interdependencies and intersections critical to implementing an effective and sustainable approach to DRI that relies heavily on RS.

1.1. Canadian RS landscape

This report identifies key stakeholders in the Canadian landscape at local, regional, provincial, national, and international levels, and considers the multiple roles they play within the key components necessary to support RS nationally.

With its RS funding program, CANARIE has been critical in advancing the Canadian RS ecosystem. CANARIE's Research Software Program funds RS development in data-/compute-intensive communities as well as non-traditional areas, accelerating research discovery by enabling access to the DRI, and promoting best practices (e.g. promoting RS reusability to avoid duplication of software (SW) reinvention). The CANARIE Research Software Portal supports access to Research Software that has been funded by CANARIE as well as contributions from the community. CANARIE hosts an annual Canadian Research Software Conference (CRSC) and is an active participant in the international RS community. CANARIE's recently launched Local Research Software Support (LRSS) initiative funds the development and maintenance of local RS teams to facilitate access to SW tools and expertise at the institutional level. Since its inception in 2007, the RS Program has awarded $50.7M in funding to Canadian RS teams, and has facilitated the focus on RS development, the establishment of RS practices, and the development of a nascent Research Software Engineer (RSE) culture and community in Canada.[4] CANARIE's more recent Research Data Management (RDM) Program invests in research projects with an RDM focus, including the provision of funding for SW development. CANARIE will transfer both RS and RDM programs to the Alliance by March 31, 2022.

The Canada Foundation for Innovation (CFI) has also provided funding focused on RS development. CFI's cyberinfrastructure challenges in 2015 and 2017 promoted the design and development of RS platforms by including Compute Canada Federation (CCF) support to enhance the research capacity involving multi-institutional consortia researchers, data scientists and SW developers. These were CFI's first forays into supporting RS platforms and were done in response to the growing awareness of the challenges of sustaining SW infrastructure. Another CFI award is the John R. Evans Leaders Fund (JELF), which is a partnership grant program that serves the needs of individual researchers in terms of research infrastructure.

Another key national DRI service provider is the CCF, which provides ARC infrastructure and services, including a unified RS environment to researchers. This national DRI framework allows for independent computing systems across all sites to install, reproduce and access the same

---

[4] All $ value in this report is in Canadian dollars, unless otherwise indicated.

collection of SW (also referred to as CCF SW repository) in a reliable and efficient manner, including over 3,000 RS applications. With more than 200 experts based at 37 partner institutions across the country, CCF provides direct support to Canadian researchers by offering consultation, expert support, and regional training on a wide range of topics. Their services cover a wide range of areas from resource allocation to data movement, and visualisation to digital humanities. In addition to the regular access and resource allocation competition for storage and computing, groups developing large-scale research software can apply to the CCF's Research Platforms and Portals (RPP) Competition, which is more focused on RS needs and requirements, providing direct support of RS and RS infrastructure for major science initiatives via deployment and development of their own platforms on CCF's ARC infrastructures. CCF's distributed ARC resources and HQP support the full range of research programs, from individual Principal Investigator (PI)-focused programs to many of Canada's major science initiatives.

Just as there is great diversity in research domains, the institutional response to supporting RS (as well as ARC and RDM) reflects a diversity defined by size, history, research focus, resources, type of efforts, and a host of other factors that make for a complex landscape. For example: colleges and institutes promote industry partnerships; national research labs implement their own local Information Technology (IT) and RS support while working with university-based collaborators; and research hospitals develop highly specialized RS that combines research with clinical practice, raising issues of security and privacy. This diversity has been accommodated in the ARC and RDM contexts through a distributed support network that has developed over the last two decades. While there is no similar distributed network in the RS landscape, a nascent community is emerging through the efforts of CANARIE (e.g. the LRSS initiative[5]), CCF, and various domain groups contributing to a coordinated approach to RS development.

Support for research programs with a RS requirement within Canada has often come through employment programs for student and recently graduated HQP from computer science, software engineering, and disciplinary fields. There is a strong history of industry-institution collaborations via co-op programs, strengthened by MITACS[6] and other government incentives. The activities of CANARIE and other organisations have built a strong foundation for the professionalization of RS development and RSEs to support the country's large SW development talent pool. Support for Indigenous data sovereignty in the 2021 federal budget[7] has RS implications, which is bolstered by Social Sciences and Humanities Research Council's[8] (SSHRC) commitment to funding research by and with Indigenous Peoples, greatly improving the prospects for Indigenous data sovereignty.

Most of the other national funding agencies focus on funding pure or applied research, and not explicitly the development of research software or infrastructure: support for RS is indirect, and typically through domain-specific research. Likewise, no funding dedicated to RS is available through provincial or regional funding agencies. Private funders and donors generally fund

---

[5] https://www.canarie.ca/program-news/research-software/

[6] https://www.mitacs.ca/en. MITACS is a national nonprofit organisation that partners with Canadian institutions to conduct research and training programs in fields related to industrial and social innovation.

[7] https://www.budget.gc.ca/2021/pdf/budget-2021-en.pdf

[8] https://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx

research in specific disciplines, rather than in RS in particular, although there are exceptions (e.g. Chan Zuckerberg Initiative, Wellcome Trust). A recent increase in collaborative programs offered by international and Canadian agencies (e.g. the joint EU Horizon 2020/CIHR[9] call) has provided funding for some RS efforts in Canada. Finally, the recent Innovation Superclusters Initiative could transform regional innovation and promote large-scale collaboration between industry and academia.

1.2. Key Challenges and Opportunities

There is insufficient support in targeted and sustainable RS funding.

Traditional research funding is fundamentally innovation-focused, leaving the need to sustain RS for the long-term up to the ability of PIs to describe their needs in an innovation context. This represents a significant gap and challenge for Canadian researchers, resulting in lost investment when RS is not maintained or generalized for reuse, and diminishes Canada's leadership role in the development of RS. A new approach to RS funding needs to recognize different types and phases of RS (e.g. experimental; emergent but production-level; established/enterprise), and devise appropriate evaluative mechanisms, metrics, and funding streams for each.

In comparison to many other countries, Canada lags behind in support of an effective, cohesive, and coordinated RS development community.

With a couple of notable exceptions, RS in Canada is typically viewed as a byproduct of domain-specific research, has lacked dedicated and sustainable funding, and only recently been recognized as a "first class" output or a core component of a national DRI strategy. By contrast, Australia and Europe have invested large sums in a focused effort to build robust RS development communities across disciplines.

The RS ecosystem is varied and complex and evolving quickly.

All SW relies on other SW components (or dependencies) for its operation, and these are diverse (e.g. operating systems, system libraries, packages, web-based software, IDEs, APIs). Writing RS for complex ARC or cloud platforms can be very challenging, as programs need to be layered to best leverage the underlying infrastructure. National services have a critical role here, whether in offering interoperable and reusable RS across platforms, or supporting education and training. RS is developed across the globe, and the rate of change in SW development and research in general is incredibly rapid: RS must evolve at a similar pace with awareness of international developments.

RS is not widely disseminated or shared and not readily discoverable, inhibiting research transparency, reproducibility, and verification.

Most RS initiatives rely on best practice commercial SW repositories such as GitHub. Given the proliferation of RS tools and platforms, scaling innovation and infrastructure to meet the range of

---

[9] EU Horizon 2020 refers to EU Framework Programme Horizon 2020; CIHR is short for Canadian Institutes of Health Research. More information: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/sc1-bhc-05-2018

requirements of diverse research communities is challenging without national coordination and support. Support for researchers to deposit their code and documentation in accessible RS repositories that provide cataloguing, indexing, preservation, and curation will be critical to ensuring that Canada's RS can become an innovative and sustainable component of the research process.

The lack of SW development skills at all stages of the education and research pipeline needs to be addressed.

There is a shortage of training in RS, with a particular need for RS training in those domains that are newer to the use of DRI in support of their research. Community-, domain-based, and regional training beyond degree programs thus remains of paramount importance. There is also a need for such sustained training as the underlying technologies in RS (and ARC) are dynamic and continually evolving. This does not apply only to fields where there has not been routine use of RS or incorporation of RS development, but also to those where there is a need to constantly upskill HQP. Education might be called out in a more unified way, including both education of RSEs and education in RS for all researchers. It is equally important to achieve greater awareness of SW sustainability best practices, which can be facilitated by organisations acting as focal points for expertise, not only to share knowledge and skills, but also to enhance international networking and collaboration.

Recognizing HQP (including researchers, students, postdocs, and research support staff) in the context of RS is key to evolving research in the digital age.

"Research software engineer" (RSE)[10] is a recently minted job title that recognizes the extent to which research efficiency and outcomes are advanced by embedding senior HQP in research teams. With the significant demand for RSEs across domains of practice, and the potential for RS developed within one domain to have applications in other disciplines, there is a need to develop a comprehensive RSE funding model. Although there are nascent RSE communities in most regions, Canada nevertheless requires focused efforts to establish both a national RSE community and a stable career path for RSEs within their host institutions and beyond. A robust recruitment and retention strategy for RS HQP, and especially RSEs, is needed.

Incentives related to metrics, funding, reward and recognition, and career progression are critical in catalyzing engagement with RS.

There is an opportunity for all stakeholders to enact policies and programs that recognize the important role RS plays in achieving research outcomes in all disciplines. Programs and incentives are needed to encourage researchers and research technical professionals (RTPs) to adhere to SW development best practices. There is an opportunity for policy makers to support initiatives that develop research assessment systems that reward SW alongside other research outputs, responding proactively when best practices are not implemented. Appropriate metrics for RS activity would facilitate the culture change needed to realize the promise of open science.

---

[10] Even though we are using this commonly accepted term in this document, in Canada RSEs are not necessarily accredited engineers.

The lack of diversity in the RS community suggests we need a more effective approach to Equity, Diversity, and Inclusion (EDI) in the RS context.

The lack of EDI in this context stems from various factors, including low enrollment of women and Black, Indigenous, and other people of color (BIPOC) in technology focused programs in higher education, and/or SW development curriculum/BS/MS/PhD programs, as well as the generally narrowing path to advancement for women and BIPOC researchers and developers. Raising awareness among stakeholders about promoting EDI in the RS profession should lead to creating and enhancing programs to expand opportunities and eliminate barriers for members of underrepresented communities interested in RS. Partnering with technical training and networking initiatives for underrepresented groups, as well as supporting EDI-oriented training programs related to RS use and development within particular research communities will help expand the talent pool. Making real headway on EDI in this area will involve challenging entrenched aspects of research culture as well as risk to individuals, so it will be crucial to create policies, systems, and workflows in research communities, teams, and within university human resources (HR) departments that protect vulnerable populations and address systemic patriarchy, colonialism, and racism.

Canada currently does not have well-developed policies, standards, and protocols to support researchers across disciplines in managing RS.

This affects researchers' ability to leverage the enormous potential of RS and diminishes Canada's competitiveness internationally. A national policy framework compatible with global practices is required to guide the sustainable development of RS in Canada.

A coordinated approach must be established to address research security and integrity and protect Canadian interests.

RS teams are typically ill-positioned to understand or to respond to the proliferation of cybersecurity threats. For instance, devising RS solutions for Science Gateways[11] that deal with health data requires specific skill sets and expertise in best practices for secure software development, as well as an understanding of data privacy policy. As concerns about cybersecurity continue to rise, it is essential to develop a robust national cybersecurity strategy beyond the scope of a single research team, service provider, or university.

There should be a balance between intellectual property (IP) management and developing a culture in which sharing information is advocated.

This culture is best fostered by individual research communities, as norms vary by discipline. The ultimate goal is for research outputs to be appropriately protected, while fostering an open and collaborative approach to research that advances Canada's world-leading research and development.

[11] https://doi.org/10.1016/j.future.2018.12.026

Federal and provincial research organisations need to develop a strategy for working more collaboratively with higher education and across governments.

The large number of users and communities, the complexity of their roles and responsibilities (international, national, regional, and local), the number of jurisdictions that are involved, and the diversity of requirements with which they must align their efforts, all highlight the challenges to, as well as the need for such collaboration. Although national coordinated leadership for RS is emerging, efforts to crystallize RS communities have been hampered by a lack of adequate funding and a formal mandate. Without coordination of investment in the context of RS, it is difficult to develop the shared policies, processes, protocols, best practices, and standards that are so essential.

1.3. Next Steps

Readiness to respond to current challenges and potential opportunities requires a thorough understanding of the current state and landscape, in agreement with short- and long- term goals and objectives. As this report summarizes the current state of the RS landscape in Canada to support a common understanding among the Alliance's members of the breadth and complexity of stakeholder engagement in this field, it also serves as a basis from which the Alliance can set a path forward for national strategy for RS in Canada. Findings and observations in this document, alongside the RDM and ARC Current State Assessment publications, are meant to help support the researcher needs assessment process, development of the Alliance's new service delivery model and funding model, and DRI strategic plan processes.

2 Introduction

As the first position paper on RS in Canada, this report surveys and summarizes RS generally as an emerging field and as an area of professionalization, nationally and internationally, and documents strengths, challenges, and opportunities within the current RS ecosystem. The purpose of this paper is to frame an understanding of the RS landscape in Canada.

Digital research infrastructure (DRI) is the collection of tools and services that allow researchers to turn data into scientific breakthroughs."[12] In February 2016, the Minister of Science, the Honourable Kirsty Duncan, met with members of the Leadership Council for Digital Research Infrastructure (LCDRI) and asked them to affirm their intention to work collaboratively to ensure that Canada has a strong DRI ecosystem in place to support the critical research undertaken by Canada's academic research community. In March 2016, the LCDRI responded to Minister Duncan's request with a letter that was signed by all LCDRI members, confirming their commitment to work together.

In November 2016, the federal department of Innovation, Science, and Economic Development (ISED) provided funding to the LCDRI to establish a short-term secretariat with full-time staffing support and tasked the group with undertaking the following:

- development of a position paper on data management (DM);
- development of a position paper on advanced research computing (ARC);
- consideration and recommendation of potential approaches for coordination.

In late August 2017, the DM[13] and ARC[14] position papers were developed by dedicated working groups for each area that consulted widely with community experts and stakeholders during their preparation. The publication of the papers was followed by further consultation.

Partly in response to these reports, through Budget 2018[15], the Government of Canada is investing $572.5 million over 5 years to enhance Canada's digital research infrastructure. The Government's goal is to provide all Canadian scientists and scholars with the digital tools they need to support Canadian innovation and occupy leading positions in their fields internationally.

---

[12] Digital research infrastructure: https://www.ic.gc.ca/eic/site/136.nsf/eng/home

[13] Data Management Position Paper For Innovation, Science, and Economic Development Canada: https://engagedrica.sharepoint.com/:b:/s/RS/Eaw-akExLoxCoAWNZJafApoBE75_I2G_XSwCoP1SJ3HLLQ?e=g6TlsQ

[14] Advanced Research Computing (ARC) Position Paper For Innovation, Science, and Economic Development Canada: https://engagedrica.sharepoint.com/:b:/s/RS/ES7QP8wWIdlFu6tHJT_NiL8BmewcMHVm7iaNxELNww4hsQ?e=WQ5JVV

[15] https://www.budget.gc.ca/2018/home-accueil-en.html

On May 6, 2019, a proposal[16] was submitted to and approved by ISED for the creation of a new organisation to coordinate funding and strategic directions for national activities related to Advanced Research Computing (ARC), Data Management (DM), and Research Software (RS).

On March 11, 2020, the Digital Research Alliance of Canada (the Alliance) was officially launched as a national not-for-profit organisation as well as a key initiative of the national DRI strategy initiated by ISED. The Alliance plays a critical role helping to advance the establishment of a researcher-focused, accountable, agile, strategic, and sustainable DRI ecosystem for Canadian researchers by providing the digital tools, services, and infrastructure. The Alliance will also coordinate and fund activities in ARC, RDM, RS and cybersecurity. One priority of the Alliance's Corporate Plan 2020-2021[17] was to prepare a position paper to assess the current state of RS, completing the review initiated with the ARC and RDM position papers. In late September 2020, a Working Group on Research Software (RSWG) was formed to develop the corresponding position paper.

The ultimate goals of the RS position paper are to:

- Build the Alliance's understanding of the current state of the RS environment
- Contribute to the definition and development of its relationship to the Alliance's service delivery model, and
- Provide a foundation for developing the strategic plan that will guide RS activities at the Alliance for the following years.

Members of the RSWG are:

- Qian Zhang, Senior Analyst, Research Software, Lead of the RSWG, Alliance;
- Susan Brown, Professor, University of Guelph;
- James Colliander, Professor, University of British Columbia;
- Brian Corrie, Technical Lead of iReceptor[18] project, Simon Fraser University;
- Gabor Fichtinger, Professor, Queen's University;
- Scott Henwood, Senior Director, Programs, CANARIE[19];
- Mark Leggott, Executive Director of Research Data Canada;
- Catherine Lovekin, Associate Professor, Mount Allison University;
- Felipe Pérez-Jvostov, Senior Analyst, Planning and Operation, Alliance;

---

[16] Digital Infrastructure Contribution Program Proposal: https://engagedri.ca/assets/documents/Digital-Research-Infrastructure-Contribution-Program-Proposal-Complete-Submisson-May-6-2019.pdf

[17] Corporate Plan 2020-2021: https://engagedri.ca/wp-content/uploads/2020/03/NDRIO-Corporate-Plan-2020-211.pdf

[18] iReceptor: http://www.ireceptor.org/

[19] https://www.canarie.ca/

- Ghilaine Roquet, Vice President Strategy and Planning, Alliance; and
- Marc-Étienne Rousseau, Director of Research Software, McGill University.

The RSWG sought feedback on the RS position paper from the broader community and stakeholder groups. We would like to thank the following groups for sharing their expertise: the Alliance's Senior Analyst team, Advanced Research Computing working group and Research Data Management working group, the Alliance's Researcher Council, and the five host sites technical leaders. A special thank you to Michelle Barker, Director of Research Software Alliance (ReSA) and Daniel S. Katz, co-founder of WSSSSPE, US-RSE and ReSA for their detailed editorial review and comments.

3 Overview of RS

3.1. What is RS?

RS is multidimensional and international. It may be defined differently according to different contexts (such as stakeholders, audiences, functions, or use cases). For this document, we define RS from a broad perspective at an emerging and complex intersection of tools, disciplines, services, platforms, hardware, resources, and the people (users, researchers, developers, stakeholders, personnel, and communities, etc.) who use and contribute to them.

RS plays multiple roles in research[20]: it is a tool that enables research (e.g. for manipulating all types of research data across the lifecycle); it can be a research result in its own right, a primary product or solution from a research project; it is also research object that is studied or used to facilitate scholarly communication throughout the research lifecycle. In this sense, RS also serves as a medium through which research ideas are shared and through which collaboration takes place. RS is critical in helping researchers make new discoveries and share their methods, data, and results with others. RS can be a research platform (a set of tools that enable research); and it can be provided as source code (that can be read and compiled), an executable (that can be run locally), and a service (that can be run remotely).

RS as a broad term may include software programs, languages, libraries, scripts, computational code, models, electronic lab notebooks, repository software, workflow management and other tools developed to support research or facilitate research processes, which themselves can be implemented in RS. RS can be packaged and shared as source code, executables, containers, services, etc.

[20] https://www.ouvrirlascience.fr/opportunity-note-encouraging-a-wider-usage-of-software-derived-from-research/

As research infrastructure, RS often refers to research platforms or services (variously referred to as virtual science labs, virtual research environments (VREs), or Science Gateways[21,22]) that instantiate a collection of RS source code, including large software frameworks, discipline-specific tools, services, and glue code, and are deployed to support both the research workflows and the communities of practice engaged in collaborative research.[23] Typically, a research platform's capabilities include data acquisition and management, processing and visualisation, storage and preservation, sharing and discovery; platforms may provide the full spectrum or a subset of components. Science Gateways may be discipline-specific and may support and enhance scientific collaboration and scholarly communication by facilitating citizen science engagement as well. Examples include data portals like Federated Research Data Repository[24] (FRDR), which allows researchers to discover, store or transfer research data, Syzygy[25], an interactive computing/learning platform developed by the Pacific Institute for the Mathematical Sciences (PIMS)[26], and the Canadian Writing Research Collaboratory (CWRC)[27], which provides a repository, tools, and a dissemination platform for literary and cultural studies. See Appendix A for a larger set of Canadian Science Gateways.

3.2. RS, Data and Research Lifecycles

Figure 1 depicts a traditional research lifecycle, starting with ideation and project planning, and cycling through acquiring and assembling resources, active project management, writing and publication, outreach, and assessment.

Whilst this overview of a research workflow looks straightforward, most stages have their own iterative sub-cycles, representing additional steps a researcher might undertake. It is also worth mentioning that the research workflow is never a sequential or linear process as knowledge is inherently iterative. Hence researchers regularly move back and forth between the steps, going from analysis to writing, and back again, before they move to the next stage of publication. RS is most often required at the discovery and analysis stage, as well as the publication stage, where archiving datasets and compiling computationally reproducible research packages are increasingly part of the process.

21 https://doi.org/10.1016/j.future.2018.12.026

22 see Figure 3 for the extent to which RS in this case serves all three DRI components

23 Scott Henwood. White Paper: Introduction to CANARIE's Research Software Program. https://www.canarie.ca/wp-content/uploads/White-Paper-CANARIE-Research-Software.pdf

24 https://www.frdr-dfdr.ca/repo/

25 https://syzygy.ca/

26 http://www.pims.math.ca/

27 https://cwrc.ca

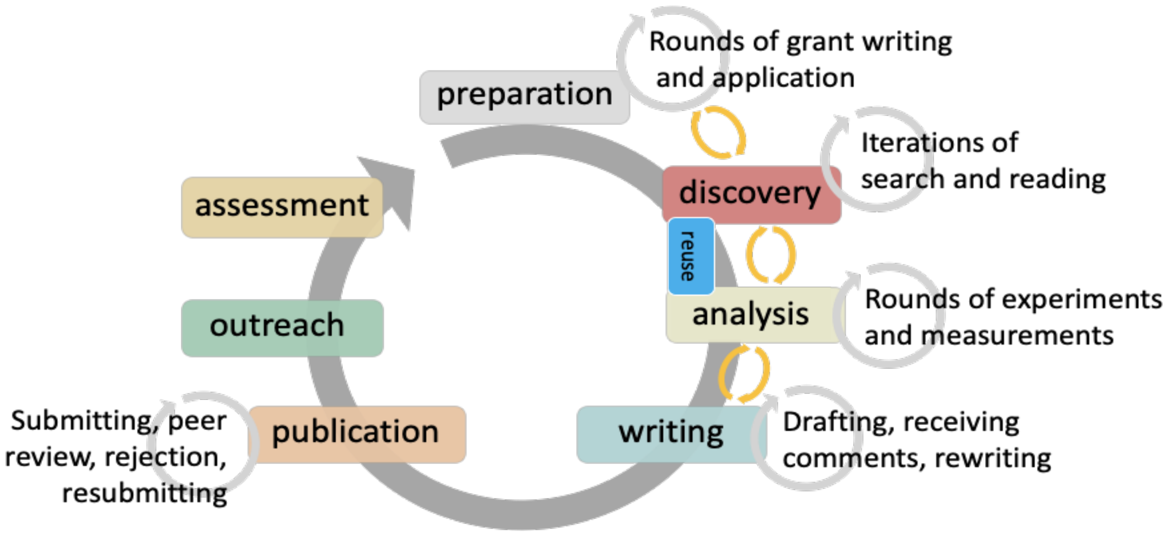

Figure 1. A model of the research lifecycle [28]

Data has special considerations in the research lifecycle, such as the decision of whether or not to preserve data, as not all data are worth preserving, or there may be special considerations (e.g. personal information) preventing them from being preserved beyond a prescribed period. Data typically have a longer lifespan than the research project, so researchers may continue to work on data after the funding has ceased, follow-up projects may re-analyse or add value to the data, and data may be reused and repurposed by other researchers, all of which rely on RS.

The RS lifecycle has some differences from the standard software development lifecycle[29], which typically goes from planning to analysis, design, implementation, testing & integration, to maintenance. The RS lifecycle is depicted in Figure 2, and includes:

- Acquiring and assembling resources (includes funding, hiring, teams and community development)
- Development
- Use and Reuse
- Recognizing contributions and impacts
- Maintenance

The development phase typically involves one or more prototyping stages. Prototyping of research software can itself, in fields such as design, human-computer interaction, and digital humanities, constitute a mode of research and a research output in itself, whether or not the RS

---

[28] Bosman, Jeroen; Kramer, Bianca (2016): 101 Innovations in scholarly communication: re-designing research support services. figshare. Presentation. https://doi.org/10.6084/m9.figshare.4054728.v1

[29] https://online.husson.edu/software-development-cycle/

is ever developed into a production system.[30,31] This means that supporting prototyping is an important feature of infrastructure. Furthermore, the prototyping stage can have substantial impacts on subsequent RS development, so participation by all members of the appropriate community of practice early in pilot programs[32] offers the opportunity to participate in ensuring that emergent infrastructure meets community needs in diverse fields. Moreover, project-based prototyping and in-house RS development can scale up into widely used infrastructure such as Jupyter notebooks[33], which are used increasingly in disciplines beyond STEM.

This model supports the definition of sustainable software considered earlier: the capacity to endure, long-term availability and reusability.

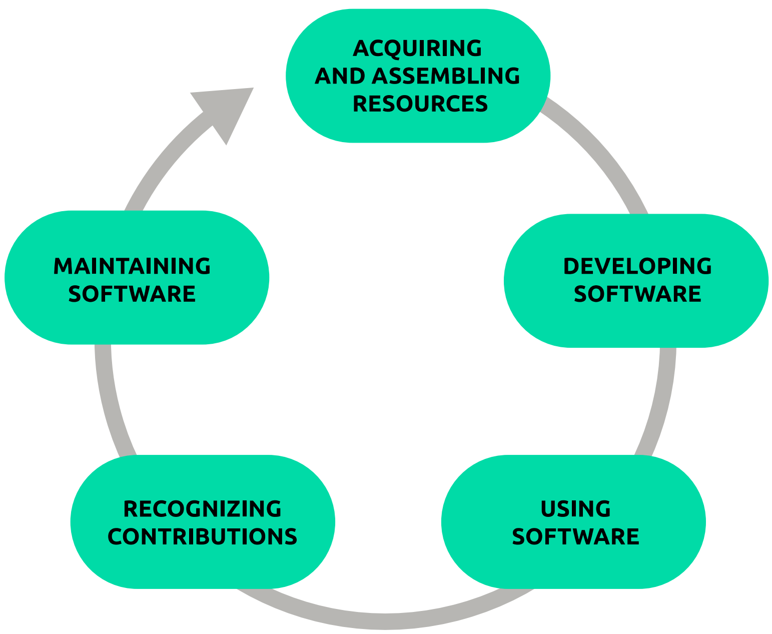

Figure 2: A model of the RS lifecycle [34]

---

[30] Galey, Alan, and Stan Ruecker. "How a prototype argues." Literary and Linguistic Computing 25.4 (2010): 405-424.

[31] Bardzell, Jeffrey, Shaowen Bardzell, and Lone Koefoed Hansen. "Immodest proposals: Research through design and knowledge." Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015.

[32] Current examples from the National Science Foundation include the Open Storage Network (https://www.openstoragenetwork.org/), the Global Research Platform (http://www.theglobalresearchplatform.net/) and FABRIC (https://fabric-testbed.net/).

[33] https://www.nature.com/articles/d41586-018-07196-1

[34] http://wssspe.researchcomputing.org.uk/about-wssspe

Both lifecycles reflect an iterative process within that specific context, but there are clearly dependencies and intersections that are critical to the deployment of an effective and sustainable approach to conducting research that relies heavily on RS. Neil Chue Hong, the founding Director and Principal Investigator of the Software Sustainability Institute, notes, "Without data it's difficult to validate results. But without code, we waste the opportunity to advance science."[35] The research and RS lifecycles are an integral and interconnected part of the research enterprise, and we have illustrated that relationship in Figure 3 below. Figure 3 highlights the central role that RS plays: overlapping with other areas of digital research infrastructure in a complex relationship to the various activities represented in Figure 1.

The activities *plan-create-process-analyze* in Figure 3 are largely the responsibility of the PIs and the research teams. When one moves to the *disseminate-preserve-reuse* phases, who's role it is to maintain RS blurs, especially when resources are limited. This may be due to the lack of funding; lack of properly trained or dedicated HQP (manifested in social sciences and humanities). The lack of any clarity around roles and responsibilities during both the operations and/or maintenance phases of development can create a precarious context for the RS to flourish. The situation gets even worse when people (especially RS developers) change roles and/or institutions (for private sectors). Another issue is RS sustainability: there are currently few options for continued support (e.g. transfer of the RS ownership) once the project's funded effort is over. In many cases, the institution's IT team is asked to assume the management of the RS. This can present huge challenges as those RS tools may not have been developed based on common practices/standards (e.g. portability, interoperability, scalability, well documentation), the IT team may not have the expertise and the capacity to maintain it for continued use. This highlights two key challenges in the RS ecosystem: ensuring best practices in RS design and development, and ensuring sufficient resources to sustain the SW.

[35] https://twitter.com/npch/status/1258388356431478784

Figure 3: The research lifecycle and RS (see Appendix B and Appendix C).

Interrelationships among Data, Compute and RS (inner circles) overlaying research lifecycle phases (outer circles). Users & Research Communities who are the agents that are engaged in the activities with RS are omitted in all these activities. Modified from the Data Management (DM) Position Paper: For Innovation, Science, and Economic Development Canada. Leadership Council for Digital Research Infrastructure (unpublished manuscript, August 31, 2017) and the Advanced Research Computing (ARC) Position Paper: For Innovation, Science, and Economic Development Canada. Leadership Council for Digital Research Infrastructure (unpublished manuscript, August 31, 2017).

3.3. Characteristics of RS

Software (SW) is pervasive in the modern digital research landscape, and integral to the creation, analysis, management and dissemination of data and publications.

RS is fundamental to research, and it should be treated as a *first-class* research output to be maintained, assessed and cited, and on an equal footing to research articles and data.[36] Features of RS that are fundamental to the Research lifecycle (Section 3.2) include:

- Metadata: administrative, descriptive, technical, structural, etc.;
- Documentation: Readme, Markdown, installation manual, etc.;
- Citability: title, creator, unique identifier, publication venue, publication date, version, type, license, language, etc.;
- Licensing status:open-source, closed-source, hybrid;
- Publication status: published, unpublished;
- Instantiation status: version, concept;
- Usability[37]: concept, work-in-progress, suspended, abandoned, active, inactive, unsupported, moved;
- Business model: free, commercial;
- Software distribution mechanism: source code[38], binary executable[39], package, container, virtual machine image, service, etc.;
- Discipline: generic, domain-specific;
- Analysis methods: quantitative, qualitative;
- Supported stages in the research lifecycle: planning, analysis, computation, visualisation, transfer, storage, publishing, curation and preservation, discovery.

SW licensing is a critical feature that helps grant clear terms for use by existing and future users, such as to use, modify, or distribute original and/or modified SW. As a public good, successful open-source projects achieve success because they have traditionally attracted a large cadre of volunteers and contributors who collaborate in designing, writing, testing, debugging, distributing, documenting, providing feedback, innovating, inspecting, and reporting SW bugs, as well as promoting the SW. This highlights a more recent shift in the nature of how SW (and in a broader sense knowledge) is created: in an open and collaborative manner, rather than in the traditional "closed-source" style.

---

[36] https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research

[37] https://www.repostatus.org/

[38] https://en.wikipedia.org/wiki/Source_code

[39] https://en.wikipedia.org/wiki/Executable

Open-source RS also presents an unprecedented opportunity for multi- and trans-disciplinary communication, and engagement with a wide range of stakeholders: an opportunity that the research community needs to understand better to leverage it more effectively. Open-source software (OSS) has also enabled research transparency by removing restrictions on access to development environments, typically accompanied by open access to the associated data. OSS has had a substantial impact on society and the economy: in addition to the broader benefits such as decreasing the total cost of ownership, enhancing security through greater scrutiny of code, and achieving vendor independence, if assigned a monetary value, the economic impact of OSS could stand at hundreds of billions of dollars per year.[40,41] The approach has spawned a massive social movement in which developers, contributors, users, researchers, practitioners, industry firms, governments, and other stakeholders collaborate to create a public good in the digital landscape.[42]

3.4. Types of RS

Inspired by the broader RS definition in Section 3.1, RS can be classified as two high-level categories below:

1. SW tools, libraries, and scripts: Components integrated at the development level that generally serve one function in the research process or infrastructure and can also be considered as foundational to more complex workflows, research platforms, and Science Gateways.

2. Research platforms or Science Gateways: These cover a broad spectrum of services, and typically present the user with sub-categories of services, components, and modules. While Science Gateways are almost always domain-specific, they may also be domain-agnostic.

---

[40] https://www.zdnet.com/article/how-much-are-open-source-developers-really-worth-hundreds-of-billions-of-dollars-say-economists/

[41] https://www.oecd.org/going-digital/mdt-roadmap-open-source-software.pdf

[42] A long-standing host for OSS projects is GitHub (https://github.com/), which as of January 2020, claimed to have more than 190 million repositories and more than 40 million users. Open-source SW has impacted global competition in the computer software (and hardware) industries where firms traditionally competed on proprietary, “closed-source” components. This has led some companies with proprietary SW products to adopt open-source solutions in their own product portfolio. It has also led to the creation of a new type of company, such as Red Hat (https://www.redhat.com/en) and MySQL (https://www.mysql.com/), which packages, distributes, and sells open-source SW and services. In many countries, government agencies have adopted explicit policies that facilitate the development and use of OSS. For example, in the fall of 2015, NASA’s Earth Science Division (ESD) published an open-source policy for the Earth Science Data Systems (ESDS) program. (https://sites.nationalacademies.org/cs/groups/ssbsite/documents/webpage/ssb_174603.pdf; https://earthdata.nasa.gov/earth-science-data-systems-program/policies/esds-open-source-policy).

By combining the typical scientific SW stack defined by Hinsen[43], Sochat's flattened taxonomy[44] and RS definition-induced classification above, the RS WG came up with a new 3-level classification of RS, which we use with this document. Under each of the categories, specific instances and examples can be further categorized into a wide array of sub-groups that demonstrates the breadth of RS applications, as follows:

1 Dedicated software created and/or used for research

1.1 Domain-specific SW packages: including libraries, scripts, tools, workflows, community modeling codes, etc. (e.g. IgBLAST[45], Astropy[46], NAMD[47], climate model CESM model[48], SPM[49], FSL[50], AFNI[51], BROCCOLI[52])

---

43 Konrad Hinsen. Dealing With Software Collapse. Computing in Science and Engineering, Institute of Electrical and Electronics Engineers, 2019, 21 (3), pp.104-108. ff10.1109/MCSE.2019.2900945ff. ffhal02117588f HAL Id: hal-02117588 https://hal.archives-ouvertes.fr/hal-02117588

44 What is Research Software? https://vsoch.github.io/2020/what-is-research-software/

45 https://www.ncbi.nlm.nih.gov/igblast/

46 https://www.astropy.org/

47 https://www.ks.uiuc.edu/Research/namd/

48 https://www.cesm.ucar.edu/models/

49 https://en.wikipedia.org/wiki/Statistical_parametric_mapping#:~:text=SPM%20is%20software%20written%20by,is%20distributed%20as%20free%20software

50 http://ftp.nmr.mgh.harvard.edu/pub/dist/freesurfer/tutorial_packages/OSX/fsl_501/doc/index.html#:~:text=FSL%20is%20a%20comprehensive%20library,is%20very%20easy%20to%20install

51 https://afni.nimh.nih.gov/

52 https://www.nitrc.org/projects/broccoli/

1.2 Domain-specific Science Gateways (e.g. FRACS[53], DataStream[54], Motus[55], Nunalit[56], OBiBa[57], WARN[58])

1.3 Domain-agnostic SW packages: including visualisation, provenance and metadata collection tools, workflow managers, integrated development environments (IDEs), etc. (e.g. NumPy[59], scikit-learn[60], YesWorkflow[61], CEDAR Workbench[62], Zotero, Kepler[63], Pegasus[64], Taverna[65], Matlab[66], Jupyter (notebooks))

1.4 Domain-agnostic Science Gateways or data repositories: (e.g. OSF, HubZero, FRDR)

---

[53] FRACS (FAIR Repository for Annotations, Corpora and Schemas) is a repository of annotations that facilitates the creation, storage, retrieval, manipulation and distribution of annotation sets on corpora of documents. https://data.crim.ca/en/

[54] DataStream is a powerful online platform for sharing information about freshwater health. https://gordonfoundation.ca/initiatives/datastream/

[55] The Motus Wildlife Tracking System (Motus) is an international collaborative research network that uses coordinated automated radio telemetry to facilitate research and education on the ecology and conservation of migratory animals. https://motus.org/

[56] The Nunaliit Atlas Framework aims to make it easy to tell stories and highlight relationships between many different forms of information from a variety of sources, using maps as a central way to connect and interact with the data. http://nunaliit.org/

[57] OBiBa is an international project committed to build OSS for epidemiological studies. https://www.obiba.org/

[58] Web-enabled Awareness Research Network (WARN) is a geohazard-detection Research Software Platform that coordinates information from dozens of sensors to notify emergency personnel and automatic safeguard systems of potentially disastrous events. https://www.canarie.ca/software/platforms/warn/

[59] https://numpy.org/

[60] https://scikit-learn.org/

[61] https://github.com/yesworkflow-org

[62] https://metadatacenter.org/

[63] https://kepler-project.org/

[64] https://pegasus.isi.edu/

[65] https://taverna.incubator.apache.org/

[66] https://www.mathworks.com/products/matlab.html

AFNI (Analysis of Functional NeuroImages) is a *domain-specific* **RS** suite of C, Python, R programs and shell scripts **to directly conduct research**, which is used analyze and display multiple MRI modalities: anatomical, functional MRI (FMRI) and diffusion weighted (DW) data.

2. General purpose software frequently used for research: Including databases, application programming interfaces or APIs, frameworks, data repositories, general data tools, search engines, analytics and visualisation tools, etc. (e.g. PostgreSQL, Neo4j, REST API, Java API, documentation tools, content management systems, Doxygen[67], Drupal[68], Excel, Grafana[69], Elasticsearch[70], Dataverse, Clowder[71], Invenio[72])

3. Software used in the process of conducting research: Including operating Systems, scheduling and task management, version control, compilers, debuggers, profilers, parallel file systems, job schedulers, system performance and monitoring tools, file editors, communication tools and platforms, file sharing or cloud storage applications and services (e.g. Linux, Windows, Jira, BBEdit[73], Photoshop, Microsoft Word, email, Slack, Twitter, Google Drive, Dropbox, git, GitHub)

For more ways to classify RS, see Appendix D.

3.5. RS in the Open Science Movement

The concept of open science means different things to different people, but ultimately encompasses: the *Philosophy* of open science which is the concept of “advancing research on the shoulders of giants”, a framing of doing research that is widely recognized; the *Policy* of open science which encourages or mandates the sharing of digital assets produced with public funds, or via a journal publisher’s requirements; and the *Practice* of open science which is where every discipline defines its own best practices and tools, including whether or not it is beneficial to share research outputs, and when, how, and what to share.

Open science “is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society.”[74] Open science extends the principles of openness to the

---

67 http://www.doxygen.nl/

68 https://www.drupal.org/

69 https://grafana.com/

70 https://www.elastic.co/

71 https://ssa.ncsa.illinois.edu/isda/software/clowder/

72 https://inveniosoftware.org/

73 https://www.barebones.com/products/bbedit/

74 https://www.fosteropenscience.eu/taxonomy/term/7

whole research lifecycle by enabling transparent and accessible knowledge, fostering sharing and collaboration as early as possible, and enhancing greater impact of scientific research, systemically transforming the way science and research are conducted[75].

Open science encompasses practices such as publishing open research, campaigning for open access, encouraging scientists to practice open notebook science, and generally making it easier to publish and communicate scientific knowledge

Open science involves various movements such as open access (OA), open monographs, open data (e.g. open data standards, open data use and reuse), open/reproducible research (open notebooks, open science workflows, open-source software), open methodology, open collaboration, open governance, open educational resources, open science evaluation (e.g. open metrics and impact, open peer review), open science policies (e.g. funders, publishers, governments, institutionals), and open science tools (such as open repositories, open services, open workflow tools), scientific social networks, citizen science[76] and research crowdfunding.

As the shift towards open science takes hold, together with the fact that most modern scholarly research relies to some extent on DRI, we should scrutinize such infrastructures to make sure that they themselves align with open principles and are available to researchers as open infrastructure (see Invest in Open Infrastructure, IOI[77]). Without these qualities, DRI is positioned more as a risk to Open Science than an opportunity. If Canadian DRI funders are committed to building open infrastructure, then we are well positioned to be a leader in the provision of a truly sustainable global scholarly infrastructure.

As noted in a CFI report[78] from 2015, many of the challenges related to the DRI ecosystem in Canada involve a shift in mindset toward greater openness and collaboration, such as a shift in thinking about "'My data'" to "'Our data,'" from "Priorities of the individual" to "Priorities of the collective," and from "A 'fragmented' DRI ecosystem" to "A 'distributed' DRI ecosystem" (CFI 2015). Although the DRI Strategy in the CFI report does not explicitly address open science, it emphasizes the need for improved infrastructure to enable greater data sharing and collaboration, and to improve the discoverability and accessibility of research data: in other words, to provide a strong foundation for open science to flourish. More recently (2020), the office of the Chief Science Advisor of Canada provided overarching principles and recommendations (including a national Open Science policy) to guide Open Science activities in Canada.[79]

The development, deployment, and maintenance of reusable RS (whether computational in nature, or that relies on any software-based analysis/interpretation) are increasingly recognised internationally as a key part of facilitating trusted, reproducible research outputs and open science. Sharing RS is a necessary, though not sufficient condition for reproducibility, and open access to RS source code helps improve the impact of the associated research. Within the context

---

75 https://www.fosteropenscience.eu/node/1420

76 https://en.wikipedia.org/wiki/Citizen_science

77 https://investinopen.org/

78 https://www.innovation.ca/sites/default/files/Funds/cyber/developing-dri-strategy-canada-en.pdf

79 https://www.ic.gc.ca/eic/site/063.nsf/vwapj/Roadmap-for-Open-Science.pdf/$file/Roadmap-for-Open-Science.pdf

of Open Science, RS can be produced as open-source with licensing that allows for modification, derivation, and redistribution (the more common approach), or it can be commercialized; both approaches are used in the research community.

3.6. Describing and Measuring Impacts of RS

3.6.1. RS Metadata

Although modern research relies heavily on SW and a growing number of researchers are engaged in developing SW as part of their research, infrastructures and best practices that support attribution, discovery, use and reuse and preservation of SW are behind that of research data and publications. Clear and complete metadata are critical to ascribing RS outputs to specific individuals and teams. Unfortunately, different SW repositories, development tools, languages and research domains denote such information in different ways, making it difficult or even impossible to work across different sources without losing valuable information. While a similar challenge exists in the research data context, the details differ substantially between the two landscapes. There is no single best practice or standard to document, index, share, archive, and manage SW, and SW citation is heterogeneous across disciplines and archives.

There are a variety of initiatives (e.g. CodeMeta[80], DataCite Metadata schema[81], DOE Code Metadata Model[82], OpenAIRE guidelines for SW metadata[83]) to facilitate the attribution of RS for different scenarios of preservation, discovery, reuse and citation that are underway. See Appendix E for details. These are generally interoperable today and are moving towards full interoperability based on a version of schema.org.

3.6.2. Licensing for RS

A SW license is the foundation of the contract between the SW creator (or copyright owner) and the end-user. The license is an agreement or instrument setting specific terms that governs how SW may be used, modified, derived, reused, or distributed by an end-user. In most jurisdictions an original work is automatically copyright protected, so it is important for the copyright owner (or the SW creator) to include a set of formal permissions and/or specific “conditions” of use granted to the licensee (or end-user). Without a license agreement, SW may be left in a state of uncertainty, as potential users may not know which limitations the creator intended. The creators may also leave themselves vulnerable to legal claims or have difficulty in controlling how their work is used. This is equally true for both proprietary SW[84] (generally commercial SW that has a fee) and free and open-source software[85] (or FOSS, SW that is made available without cost). Although end users may balk at overly restrictive SW licenses, the uncertainty caused when no

---

80 https://codemeta.github.io/

81 https://schema.datacite.org/

82 https://github.com/doecode/software-metadata/blob/master/XMLSchema/MetadataSchema.xsd

83 https://software-guidelines.readthedocs.io/en/latest/

84 https://en.wikipedia.org/wiki/Proprietary_software

85 https://en.wikipedia.org/wiki/Free_and_open-source_software

license is given can also discourage those who wish to make use of the work. For the case where the open-source software was created by multiple contributors, the management of ownership and licensing of copyrights gets even more complicated. Often the local institutional policies or union contracts will clarify the ownership.

Another way of managing and further protecting the rights associated with contributions that is often overlooked is through a Contribution License Agreement (CLA)[86], which could be customized to expressly define, describe and grant a copyright license of the software contribution, rights and obligations of the contributing authors, the open-source project, and/or the project's maintainer, if there is any. For the long term, CLAs not only help maintain the integrity of the project but also provide legal assurances for each of the project's participants, facilitating dispute resolution regarding licensing or ownership of the code base contributions. CLAs may include provisions that address copyright infringement or alternative dispute resolution, or even permit changing licenses over time, or distributing the contribution simultaneously under separate (proprietary) licenses without having to seek every single contributor's approval in advance. CLAs are not standardized, so contributions to different open-source projects may be governed by multiple CLAs. Although CLAs may not be ideal for every project such as small open-source projects, larger projects may require formal CLA agreements from their contributors. Specific implementation choices for a CLA will vary from a short and simple CLA agreement to a more detailed legal instrument.[87]

In academia, it is worth noting that licenses can be used to facilitate access to RS as well as restrict it, so it is important to choose the right SW license. While Creative Commons licenses (including CC-BY) are commonly used in the scholarly context, they are not recommended for use with software[88], as SW-specific licenses (e.g. Apache-2.0[89], MIT[90], GNU0GPLv3[91]) provide a much clearer guidance for potential users. It is also worth noting that under the policies (ie. union contracts) of most academic institutions in Canada, it is typically the creator who owns full intellectual property[92] (IP) rights to their works, while the situation in colleges and other types of research institutions typically has the IP rights owned by the institution or funding partner. One must also bear in mind that each funding source (both public and private), can have their own specific requirements when it comes to IP, and they may or may not be negotiable.

---

[86] https://www.finnegan.com/en/insights/articles/what-you-should-know-about-contributor-license-agreements-in-open-source-projects.html

[87] Some well-known open-source project examples are Apache Software Foundation (https://www.apache.org/), Django (https://www.djangoproject.com/) and Eclipse Foundation (http://www.eclipse.org/projects/). The Islandora project is governed by a Canadian Foundation, and includes different forms of CLAs (https://www.islandora.ca/contribute), including Individual Contributor License Agreement (https://islandora.ca/sites/default/files/islandora_cla.pdf) or Corporate Contributor License Agreement (https://islandora.ca/sites/default/files/islandora_ccla.pdf).

[88] https://creativecommons.org/faq/#can-i-apply-a-creative-commons-license-to-software

[89] https://opensource.org/licenses/Apache-2.0

[90] https://opensource.org/licenses/MIT

[91] https://choosealicense.com/licenses/gpl-3.0/

[92] https://en.wikipedia.org/wiki/Intellectual_property

Software license terms are diverse, but two common categories are used across the spectrum of SW licensing strategies: proprietary; free and open-source. In recent years, a third category has emerged: a hybrid of the two.

Proprietary

This type of SW is usually closed-source (aka proprietary), which includes any variation that does not allow the source code to be viewed and reused. As a result, the SW under proprietary licenses are typically distributed only in binary form and examination of the source code or reverse engineering of any part of the code is typically not permitted. The primary purpose of a proprietary SW license is to limit the use of SW according to the owner's business strategy. That is why proprietary licenses are often restrictive for end users. For example, they typically allow the SW to be used and installed on a single computer, along with additional limitations such as limited user accounts, inability to reuse code or to redistribute or alter the product. In academic settings, proprietary SW may occasionally release source code "for inspection purposes only" due to scientific publishing and/or peer-review requirements. Source code release may also be part of purchase agreements (e.g. an escrow clause), in the event of a company going out of business, or ceasing support for a specific application. In some cases, software developed in research projects, even with public funding, can become proprietary as the chosen method from the developers and funder to meet the need for a reliable source of funding to sustain it. Sometimes in these cases, there is an agreement for free or low-cost distribution to academic institutions, while other users are charged a higher fee.

Free and Open-Source Software (FOSS)

FOSS represents a fundamentally different approach to SW licensing, in which the source code is made available to users, and includes the ability to use, view, distribute, modify, and reuse. The primary advantage of FOSS is how it facilitates access and maximizes openness, thus minimizing barriers to SW use, dissemination and follow-up derivation and innovation, ultimately leading to better SW. Some other benefits of a FOSS strategy include: wider adoption, greater security, robust and sustainable code[93] with quick bug identification and fixing, community building, ease of collaboration and user contributions. Source code availability and access is increasingly important in the context of scientific research, where peer review, reproducibility, and building upon prior work are integral to the advancement of science. Additionally, FOSS licenses can extend the lifespan of a piece of SW (i.e. sustainability) beyond the direct involvement of the original creators by simplifying continuous development and adoption by successors.

There are a wide variety of popular FOSS licenses. Information on how to choose an open-source license can be found at ChooseALicense.com, SPDX License List[94], REUSE project[95] and from the Open Source Initiative[96]. The National Initiatives for Open Science in Europe (NI4OS Europe)

---

[93] https://openresearchsoftware.metajnl.com/article/10.5334/jors.118/

[94] https://spdx.org/licenses/

[95] https://reuse.software/

[96] https://opensource.org/licenses/alphabetical

released the License Clearance Tool (LCT)[97], a web portal that helps provide a guided approach for establishing the appropriate open-source licence for the creation of new research outputs (e.g. dataset, media, software etc.), or for the reuse of existing unlicensed content. Some FOSS licenses can also be formulated to permit the derivation of closed-source alternatives, which could be considered one path to Hybrid licenses.

Hybrid SW

This licensing model is also called dual- or multi-licensing[98]. As the name suggests, the hybrid approach combines a FOSS license with a proprietary closed license, facilitating the practice of distributing SW under two or more sets of terms and/or conditions. This occurs when neither one of the other two licensing models meet the SW developers' needs. Under this hybrid licensing scheme, the rights owner can choose which license to be applied on a case-by-case basis. The downside of this approach can be a significant extra burden for the rights owner to apply, administer, and enforce multiple licenses, which will generally limit the wide adoption of hybrid license models to large SW development initiatives.

3.6.3. RS and Research Reproducibility

The main purpose of research reproducibility is to provide enough information about a study such that the research methods used can be utilized by another researcher (method reproducibility) to generate the same outcomes (results reproducibility) given the same materials/inputs[99]. Research reproducibility is a critical and continuous component of the scholarly communications process as articulated in the FAIR principles for making data findable, accessible, interoperable, and reusable. RS, together with the associated research outputs, promotes knowledge production, sharing and mobilization by enabling reproducibility. A best practice approach to reproducibility advances the case for both open science and digital stewardship via open access, open data, and ideally open source code. Software versioning, a robust testing/quality framework (e.g. verification and validation), code repositories, and portability to Cloud have all helped to drive the rapid evolution of research reproducibility.

Since the wide adoption of FAIR principles on data, many initiatives have emerged to apply the FAIR principles to other researcher artefacts such as FAIR for training materials[100], FAIR Semantics[101], and more broadly FAIR Services[102] and FAIR Digital Objects[103]. Some

---

[97] https://lct.ni4os.eu/lct/login

[98] https://en.wikipedia.org/wiki/Multi-licensing

[99] What does research reproducibility mean? Science Translational Medicine, 01 Jun 2016: Vol. 8, Issue 341, pp. 341ps12, DOI: 10.1126/scitranslmed.aaf5027

[100] https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007854

[101] https://www.fairsfair.eu/events/recommendations-fair-semantics

[102] https://www.scilit.net/article/b22490fab3437063a08bb94ef6eb9b82

[103] https://datascience.codata.org/articles/10.5334/dsj-2020-015/

work[104,105,106] has been extended to the application of RS and the guiding principles that translate this purpose into RS context[107] are still underway. Specifically, initiatives like FAIR for Research Software (FAIR4RS) WG[108] and CURE-FAIR WG[109] have been working on coordinating a range of community-led discussions on how to define and effectively apply FAIR principles to RS.

More information regarding best practices for RS management and computational reproducibility is provided in Appendix F.

3.6.3.1. RS Sharing

The availability of RS is critical to reproducible research. Journal publishers and other research stakeholders (such as funders like the Wellcome Trust[110]) have recently begun to explicitly include RS in policies pertaining to the management and sharing of scholarly research outputs, mandating that RS as well as data be made openly available for review and reproducibility[111]. The general lack of formal mandates or requirements for doing so, along with the funding needed to provide quality outputs, can make it easy to ignore. What complicates matters further is that the RS expertise and infrastructure that are available to researchers can be inconsistent and heterogeneous, and many researchers are not professional SW developers with formal training in best practices. Many researchers build and develop SW not for others, but for their personal research, and thus may tend towards fast-return coding efforts that solve their immediate problems, particularly in humanities and social sciences (HSS) fields where solo research is still the norm and larger team-based funding is scarce. There is also pressure to publish or disseminate a result as soon as possible, advancing in the "publish or perish" race. Sharing software can be seen as giving competitors a leg up. An unintended consequence is that the result may not be reproducible or reusable by others, compromising peer review and

---

[104] Chue Hong, Neil; Katz, Daniel S. (2018): FAIR enough? Can we (already) benefit from applying the FAIR data principles to software?. figshare. Presentation. https://doi.org/10.6084/m9.figshare.7449239.v2

[105] Christopher Erdmann, Natasha Simons, Reid Otsuji, Stephanie Labou, Ryan Johnson, Guilherme Castelao, … Tim Dennis. (2019, February). Top 10 FAIR Data & Software Things. Zenodo. http://doi.org/10.5281/zenodo.2555498

[106] Aerts, Dr. (PhD) P.J.C. (DANS) (2017): Sustainable Software Sustainability - Workshop report. DANS. https://doi.org/10.17026/dans-xfe-rn2w

[107] Lamprecht, Anna-Lena et al. 'Towards FAIR Principles for Research Software'. 1 Jan. 2020 : 37 – 59. https://doi.org/10.3233/DS-190026

[108] https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg. The final draft of the FAIR principles for RS are in the final stage of formal community review - https://www.rd-alliance.org/group/fair-research-software-fair4rs-wg/outcomes/fair-principles-research-software-fair4rs

[109] https://www.rd-alliance.org/groups/cure-fair-wg

[110] https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

[111] Stodden V, Guo P, Ma Z (2013) Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. PLoS ONE 8(6): e67111. https://doi.org/10.1371/journal.pone.0067111

advancement, and leading to the replication of efforts and waste of funding across multiple projects.

Another disincentive is the concern that can stem from scrutiny in the peer review process. It is generally agreed that open code leads to better SW, provided a community of contributors are willing to improve or comment on it. GitHub[112] is one example of a code sharing platform that has features specifically designed for the scientific community to support effective code collaboration and sharing[113]. Code sharing is sometimes considered a 'double-edged sword' to some researchers, since the more it is shared, the higher the risk that mistakes and bugs will surface. While in and of itself such exposure is beneficial for science in general, research can be highly competitive, and there is concern that the open review process may cause disproportionately harsh criticism and even jeopardize one's reputation[114]. In a system where careers advance through publication, the fear of misinterpretation or misuse of code, or discovery of errors, thereby compromising results, can be a big concern. With sufficient training and access to tools like SW testing frameworks, the community should be able to help mitigate this challenge. With the proper licenses and attribution, as well as regular use of SW citations that will facilitate the culture change needed, SW sharing concerns should diminish. The emergence of software bill of materials (SBOM) is a useful addition to RS projects/code.[115] With a robust approach to reproducibility, the researchers, scientific findings, and datasets all receive greater recognition and credibility, further strengthening the record and impact of scholarship as well as an individual's productivity and competitiveness.

3.6.3.2. RS Access

The Accessibility principle (5) of the FORCE11 Software Citation Principles states that "software citations should permit and facilitate access to the software itself and to its associated metadata, documentation, data, and other materials necessary for both humans and machines to make informed use of the referenced software." While this does not require that the RS be freely available, the metadata should be, and should provide sufficient information for the RS to be accessed and used. For commercial software, the metadata can still provide information that permits and facilitates the access to the specific software, which may be a company's product number or a link to a landing page with information that allows the software to be used for free for academic review purposes or be purchased. However, considering accessibility to be achieved through mere retrievability is not enough: a description of how to access a working (e.g. binary) version of the RS should be made available along with sufficient documentation and licensing information. Furthermore, metadata about how to run the RS, including dependencies such as access to research datasets, other software dependencies, and runtime environments should be provided for RS to be truly accessible.

---

112 https://github.com/

113 https://github.blog/2014-05-14-improving-github-for-science/

114 Cope, J. (2015). Why not to share your data: The "crackpot" argument. eRambler Blog. Retrieved from http://erambler.co.uk/blog/crackpot-argument-against-data-sharing/

115 https://fossa.com/blog/software-bill-of-materials-formats-use-cases-tools/

3.6.3.3. RS Portability

The concept of software portability in this document is defined as the ability to move a RS application between different environments, such as platforms or machines, and so that the application produces the same results given the same inputs. Due to the wide diversity of software applications and dependencies, this level of portability may have different meanings to different people[116]. In one scenario, the software is portable only if the executable files can be run or reinstalled on a new platform without change: binary portability. In a more common scenario, the source code is made available, and the executable program can be built from the code for reuse in an appropriate target environment. It should be noted that this scenario involves both the access to and transportation of code, and any conversion which may be required (e.g. modification of the original code to accommodate specific machine architectures, environments, or local research team needs). A third scenario is one of intermediate-level portability, which entails aspects of both previous scenarios to achieve the desired outcome. Generally, one can assume that the cost to transport and adapt the software to a new environment is less than the cost of redevelopment, although this is not always the case. Once an RS application is ported to multiple environments, more users will benefit. Portability is a key issue for development cost reduction so that developers can spend more time on new efforts.

As portability is recognized as a desirable aspect of software quality, more attention is being paid to increasing reuse and research reproducibility. It may also be argued that research software portability is itself a response to reuse and reproducibility. While the primary entity to be ported is usually a complete software system or application, in some cases auxiliary software components such as dependencies, interfaces, tools, libraries or databases may need to be ported as well. Production of complete documentation for the porting process should be viewed as a reproducibility activity, and one that facilitates the development of more qualified and experienced RS teams.

When it comes to compute-intensive research, such computational environments can evolve very rapidly, with many different versions of the same operating systems and software components in concurrent use. For a project that is dependent on specific pieces of SW from diverse sources, it is not uncommon for components to change, sometimes frequently. A wide range of changes (e.g. updates to low-level software, a minor configuration change, changes in operating systems) may lead to unexpected results, and may cause the entire computational workflow to fail[117] or simply produce slightly or dramatically different outputs. One solution is to provide both an identical copy of the operating system and exact versions of all SW dependencies for others to recreate the original computational environment. Together with good documentation researchers can more easily create new research environments and workflows, and vastly improve transparency, code portability, reuse, and reproducibility. However, using many of these environments that freeze code in some state has to be balanced with desirable changes such as bug fixes and the additional of new features. Examples of systems currently being used by

---

[116] Mooney, J. D. (2004). Developing portable software. In Information Technology (pp. 55-84). Springer, Boston, MA.

[117] Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., et al. (2015). Reproducibility of neuroimaging analyses across operating systems. Front. Neuroinform. 9:12. https://doi.org/10.3389/fninf.2015.00012

researchers to capture complete research environments include virtual machines (e.g. Oracle's VirtualBox[118]) and containers (e.g. Docker[119], Podman[120], Singularity[121], Sarus[122], NERSC Shifter[123], HPC Container Maker (HPCCM)[124]). Vagrantfiles[125] and Dockerfiles[126] are common examples of machine-readable plain-text instructions for making virtual machines to an exact specification.

3.6.3.4. RS Reuse

SW reuse "entails capitalizing on existing software and systems to create new products"[127], and has been a research topic for more than 50 years, since its inception at the first International Conference on Software Engineering (ICSE) in 1968[128]. Researchers and practitioners continue to develop and enhance tools, methodologies, processes, practices, and techniques for increasing productivity, quality, and cost savings. A similar definition considers research innovation itself: "RS reuse entails capitalizing on existing software and systems for other applications or to create new *research*." Under this broad definition, several contexts coexist. For example, there are two main approaches to RS development: *with-reuse* (development using pre-existing components) and *for-reuse* (development of new reusable components for research, such as software libraries)[129]. RS reuse can also be considered in a domain context: *vertical-reuse* (reuse of RS in a given application domain) and *horizontal-reuse* (reuse of RS components in *other* applications or across more than one application domain)[130]. Furthermore, RS reuse can

---

[118] https://www.virtualbox.org/

[119] https://www.docker.com/

[120] https://podman.io/

[121] https://sylabs.io/singularity/

[122] https://user.cscs.ch/tools/containers/sarus/

[123] https://www.nersc.gov/research-and-development/user-defined-images/

[124] https://github.com/NVIDIA/hpc-container-maker

[125] https://www.vagrantup.com/docs/vagrantfile

[126] https://docs.docker.com/engine/reference/builder/

[127] IEEE Standard, 1517-2010 - IEEE standard for information technology–system and software lifecycle processes–reuse processes, 2010.

[128] McIlroy, M. D., Buxton, J., Naur, P., & Randell, B. (1968, October). Mass-produced software components. In Proceedings of the 1st international conference on software engineering, Garmisch-Partenkirchen, Germany (pp. 88-98).

[129] De Almeida, E. S., Alvaro, A., Lucrédio, D., Garcia, V. C., & de Lemos Meira, S. R. (2005, August). A survey on software reuse processes. In *IRI-2005 IEEE International Conference on Information Reuse and Integration, Conf, 2005.* (pp. 66-71). IEEE.

[130] Martin Griss, Ivar Jacobson, Chris Jette, Bob Kessler, and Doug Lea. 1995. Systematic software reuse (panel): objects and frameworks are not enough. SIGSOFT Softw. Eng. Notes 20, SI (Aug. 1995), 17–20. DOI:https://doi.org/10.1145/223427.213969

just describe maintaining and updating an old piece of RS to a new and advanced version (of prior *used* systems).

With a focus on RS access and portability, RS reuse will enhance transdisciplinary research productivity in various ways. The primary advantage is the savings in time and resources, reducing redundant work required to create RS with similar functionalities from scratch. Moreover with RS reuse, researchers are able to spend more time undertaking new research by leveraging existing assets: modifying, updating, adapting and advancing existing RS for their own research purposes (while giving appropriate credit to the original code authors). RS reuse will also enhance the reliability, productivity, and quality of RS, and promote community building by connecting people with similar research interests, thus creating a collaborative environment for idea exchange and motivation.

The extent to which RS can be reused is contingent on various factors, including the technical quality that is crucial for motivating others to reuse the RS, along with the quality of documentation, implementation details, level of portability (e.g. containerization) and full access to the related data. This leads to the quality-conscious mindset (e.g. testing, reviews, continuous development, version control) of code authors as a prerequisite for achieving the desired high quality RS product. On the other hand, effective RS reuse can be achieved via a supporting infrastructure (e.g. registries or repositories) for publishing, searching, integrating and cross-referencing the RS and related assets. The culture and human factors (e.g. experience, skill sets, education, incentives and motivation, etc.) matter, too.

It is also worth noting that various workflow languages have been developed to help researchers document, share, modify, and re-run experiments, providing a critical tool for effective RS reuse. There are approximately 280 workflow management systems - the most popular ones being developed for science are adding support for Common Workflow Language (CWL) and CWLProv. CWL has substantial uptake in science VREs, especially in biomedical platforms, but there are also applications in hydrology, radio astronomy and high energy physics. CWL supports best practices, for example the I/O files all have a PID, and rich descriptors (metadata); CWL also supports Docker and CVMFS, both of which are prevalent packaging systems found across science infrastructures.

3.6.3.5. RS Citation

RS is a critical part of the digital research process across the research enterprise, as integral as a research paper or dataset in terms of facilitating the dissemination and understanding of research. It is important that RS is properly cited because RS citation allows researchers and/or RS creators to publish RS rather than, or in addition to, papers and data as a way to participate in the scholarly system and its reward system. The scholarly communication system that underpins research relies on accurate citation mechanisms to ensure research integrity, give appropriate credit, and represent relationships with other research products. As one of the steps towards scientific reproducibility, researchers need to be able to uniquely identify and access all inputs, including the specific version of RS packages that are used to produce the original results, contributing to the transparency and traceability of research results.

Increasingly sophisticated systems are being built to respond to the need for best practice tools for code hosting and collaboration, automating and simplifying the discovery and dissemination process, facilitating queries across scholarly records, and ensuring interoperability. For example, RS collaboration and dissemination platforms include GitHub and Bitbucket, which are also for collaborative development. IPOL is a publishing platform for image processing research. For versions of RS package distribution, existing platforms and/or registries include PyPI for Python, CRAN for R, Awesome Research Software Registries, Astrophysics Source Code Library (ASCL) for astrophysics, MLOSS, DLHub and OpenML for machine learning open-source software, Madagascar and Knowledge Exchange. Software Heritage is an open, non-profit initiative and infrastructure for long-term software preservation and archiving. Existing reproducible platforms/initiatives include ResearchCompendia, nanoHUB, RunMyCode, Code Ocean and Whole Tale, etc.

The scholarly citation infrastructure has developed alongside scholarly publishing of books and journal articles, and thus there has been uneven support across the scholarly ecosystem for acknowledgement and citation of RS. However, there is growing acknowledgment in all disciplines that RS should be cited along with other research outputs, though the approaches vary widely by discipline as well as by publishers.[131] Citing RS helps further research, and provides the means for other researchers to access RS in order to:

- enhance research transparency and thus the understanding of research outcomes;
- enable peer review, verification and reproducibility of research findings;
- foster collaboration and reuse;
- minimize duplication of effort;
- encourage building on others' work; and
- elevate RS to the level of a "first-class" research object in the DRI ecosystem.

The FORCE11 Software Citation Working Group developed and recommended a set of six citation principles to encourage broad adoption of a consistent approach to SW citation: importance; credit & attribution; unique identification; persistence; accessibility; specificity. The Software Citation Checklist for Developers[132] and Software Citation Checklist for Authors[133] provide even more practical information for those seeking to improve their practice. The RDA/FORCE11 Software Source Code Identification WG[134] provided an overview of the current state-of-the-art of persistent identifiers (PIDs) for software. After analyzing use cases and identifier schemes from different academic domains and industry, the WG concluded that, rather than a one-size-fits-all PID, a strategy of combining multiple PIDs, both extrinsic (created by a service external to the development environment, such as a DOI) and intrinsic (such as a hash created by the development environment), is needed to capture the fundamental RS use cases (discoverability, access, persistence, reproducibility and reuse).

---

131 https://doi.org/10.7717/peerj-cs.86

132 https://doi.org/10.5281/zenodo.3482769

133 https://doi.org/10.5281/zenodo.3479199

134 https://www.rd-alliance.org/group/rdaforce11-software-source-code-identification-wg/outcomes/use-cases-and-identifier-schemes

3.6.3.6. RS Curation

In order to validate scientific research output, RS needs to remain accessible and usable over time; and curating and preserving RS will assist in the integrity and transparency of modern research endeavors. Like other digital objects, RS requires active and ongoing management through its lifecycle, and this can be defined by two primary activities: curation activities that enable data discovery and retrieval, maintain its quality, add value, and provide for reuse over time[135]; and SW maintenance activities, which ensure that the SW remains functional over time. The working definition of RS curation is: *"Software curation encompasses the active practices related to the creation, acquisition, appraisal and selection, description, transformation, preservation, storage, and dissemination/access/reuse of software over short- and long- periods of time."*[136] In this sense, curating and preserving RS over time will help facilitate the long-term access and use of RS. SW maintenance requires a different skill set more common in the developer community, and is focused on ensuring that the appropriate technologies and frameworks (e.g. containers with all associated code) are maintained, tested, and accessible.

As an emerging set of practices, RS curation may vary, so identifying the needs and activities from defined use cases is crucial. Further exploring the types of RS, and understanding the multiple stakeholders, uses, functionalities and even the location of RS can help contextualize curation efforts. For example, faculty members and students may create RS as part of coursework or projects, use or reuse RS in the process of conducting new research: these are very different use cases that may call for different curation approaches. Project-specific websites, and cyberinfrastructure are increasingly common, and collaborative projects push disciplinary boundaries and practices. All these scenarios reflect the complex roles RS can play in the scholarly enterprise. As a use case, consider the process of writing a piece of RS as a tool for data analysis. This simple example demonstrates how RS can simultaneously serve as both a research output that needs to be preserved, and as a methodological means to an analysis result.

RS curation and preservation is a field of study and practice dedicated to ensuring continued access to digital artifacts of enduring value and interest, but also in fact a complex mix of policies, practices, and infrastructure to support the long-term maintenance of access to these artefacts beyond media failure or technological change. A common approach in digital curation and preservation is to characterize the significant properties or "essence" of the digital object[137]. What are the essential components of RS as a scholarly object to be curated and preserved? Defining the boundaries of RS is difficult, partially because RS has diverse representations across contexts and disciplines (see the section What is RS). Unlike research datasets and publications, RS is executable, highly iterative, and often dependent on multiple dynamic elements that are sensitive to changes and difficult to track. Additionally, RS is potentially always evolving: a few lines of added code may impact multiple dependencies; the deprecation of SW and/or SW libraries is a common occurrence in SW development. The challenges of RS curation are not just limited to providing access and ensuring the bits are preserved. Enabling adequate use and reuse of RS involves technical interventions across the lifecycle to facilitate different preservation strategies

---

135 http://hdl.handle.net/2142/3493

136 https://drmaltman.wordpress.com/2017/01/21/guest-post-alex-chassanoff-on-software-curation/

137 Heslop, H., Wilson, A., & Davis, S. (2002). National archives green paper: an approach to the preservation of digital records. NAA.

(e.g. execution, migration, emulation, etc.). The wide range of robust frameworks and approaches for curating and preserving RS as a complex digital object represents a significant challenge for sustainable access, thereby hindering research reproducibility.

3.6.3.7. RS Impact and Value

Software is clearly fundamental to research, but how much do we rely on SW? An online survey "How do scientists develop and use scientific software"[138] conducted in October–December 2008 received almost 2,000 responses: 91% indicated using software is important for their own research; 84% that developing SW is important for their own research. The UK Research Software Survey 2014[139] showed that 70% of researchers responded that their research would be impossible without software. Three years later, the 2017 US National Postdoctoral Association Survey found a similar result (63%)[140]. Many people in the research community are developing their own code, highlighting the impact of RS in the research enterprise.

The singular focus on journal impact to assess researcher's worth and productivity is beginning to include the recognition of SW development and other activities as part of academic reward and promotion guidelines. For example, the Declaration on Research Assessment (DORA) has the goal to improve the ways in which the outputs of scholarly research are evaluated, and has been signed by over 20,000 individuals and organisations in 148 countries around the world.[141,142,143] Similar to data and publications, scholarly impact factors and metrics for RS may include the number of views, downloads, citations, user ratings (e.g. GitHub stars), contributions to a project or repository (e.g. forks), social media mentions, etc. The major benefits of effective RS management, sharing and reuse, and generally measurable via specific metrics are highlighted below.

RS facilitates the development of sustainable networks, such as the emergence of the Research Software Engineers (RSEs). The International RSE community began with UK-RSE[144], which was then enlarged with a number of newly formed international RSE associations[145]. As HQP, RSEs can move from one project to another, from one discipline to another, gaining additional

---

[138] J.E. Hannay et al., "How Do Scientists Develop and Use Scientific Software?" *Proc. ICSE Workshop Software Eng. for Computational Science and Eng.*, 2009, pp. 1–8.

[139] Simon Hettrick. (2018, February 23). softwaresaved/software_in_research_survey_2014: Software in research survey (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.1183562

[140] https://doi.org/10.6084/m9.figshare.5328442.v3

[141] https://sfdora.org/signers/

[142] https://www.scienceguide.nl/2019/11/dutch-universities-and-research-funders-move-away-from-the-impact-factor/

[143] http://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaquette_ScienceOuverte.pdf#page=7

[144] https://society-rse.org/

[145] More recently formed and forming international RSE associations including de-RSE (http://de-rse.org/), nl-RSE (http://nl-rse.org/), nordic-RSE (http://nordic-rse.org/), us-RSE (http://us-rse.org/), RSE-aunz (https://twitter.com/rse_aunz), be-RSE (https://www.be-rse.org/), and the DHTech Special Interest Group of the Alliance of Digital Humanities Organizations (https://dh-tech.github.io/#/).

experience, and enhancing the RSE community at large. RSEs can be exposed to a wide range of disciplines when part of an institutional RSE team or have a focus on a particular discipline when part of a PI's team. This range of RS and domain expertise is a strength of a community that is increasingly involved in cross-disciplinary research. CANARIE's Canadian Research Software Conference (CRSC)[146] is a good example of an event that brings the RS community together from across the country to meet and network fostering future transdisciplinary collaborations.

Effective SW sharing and reuse can potentially lead to RS developers and creators gaining recognition. SW citation enables recognition of scholarly effort, thereby facilitating career advancement and increased professional reputation.

As an example of the impact of RS consider OpenPNM[147], developed by the Porous Materials Engineering and Analysis Lab (PMEAL) Group[148] at the University of Waterloo, an open-source software package providing porous media researchers with a ready-made framework for performing a wide range of pore network simulations. The PMEAL Group has engaged in several research collaborations since its release, many not otherwise likely to have materialized. In another example, the McGill Centre for Integrative Neuroscience (MCIN) is emerging as an international leader by actively developing and providing computationally-intensive technologies (including software tools and platforms) that are openly available for advancing collaborative brain research and clinical care across Canada and international partners, as well as for use by Canadian and international research communities.[149]

These and other examples in Canada show that RS can have a substantial reputational impact in internal and external communities of practice, nationally and internationally.

Researchers strive for reliable and defensible results, which are more readily achieved using robust, well-tested RS. Well-documented and accessible SW makes it easier for others to reproduce and verify results. It is in Canada's best interests to protect the large annual research investment by ensuring that the research community has the skills needed to create reusable software.

Developing SW can be very expensive, but there are times when a specific research problem requires a "from scratch" or greenfield development effort. However, in many cases, reusing RS can save researchers time by leveraging code that already exists, and resolving problems that have been solved by others. Reusable SW can be easy to adopt and extend, assuming the right combination of skills are represented in a research team, and the reused code is well designed and implemented. By building on existing RS, researchers can invest more time into actual research, and save on scarce resources, which can then be invested into new research rather than software development. To ensure a robust and effective research ecosystem, we need both new RS developments and RS reuse to foster innovation and efficiency: reuse is about applying

---

[146] https://www.canarie.ca/software/canadian-research-software-conference/

[147] http://openpnm.org/

[148] http://www.pmeal.com/

[149] https://mcin.ca/technology/

known techniques to gain understanding, while development is about creating new techniques to gain understanding.

Using a common pool of best practice RS, regardless of the discipline, can speed the review and acceptance of research results. This is apparent both in the ease of access to the RS tools used in specific research projects, as well as the increased trust the larger research community has in the accuracy and efficacy of the associated RS.

Many types of research data can be meaningless without SW to read and interpret it. Well-designed and sustainable SW helps to ensure long-term access, use and reuse of research data, leveraging reproducibility and providing the best return on the investment made in producing the data.

Code sharing and reuse facilitate the adoption of standards and establish communities of practice that can provide better peer support, mobility of personnel between projects, and so on. Code sharing and reuse becomes one of the ways in which research communities are created and sustained. RS can also have non-scholarly impacts such as economic benefit and social good: new job opportunities; RS as end products with support and associated services; other forms of intellectual property (e.g. patents) that become part of the innovation pipeline. Some domain-specific RS can have a profound impact on industry and economy, such as in medical diagnosis, engineering techniques, simulations (e.g. weather, engineering, materials, epidemiology, chemical process), or communication and collaboration. As a more detailed and current example through the lens of COVID-19, see Appendix G.

3.6.4. Long-Term RS Maintenance and Sustainability

Software sustainability is key to reproducible science, as it provides a critical tool for the effective review and analysis of published results, which can of course lead to new research efforts. The Software Sustainability Institute[150] defines sustainability as ‘the software you use today will be available - and continue to be improved and supported - in the future’. Inspired by Daniel S. Katz’s blog[151] and related literature[152], the authors think of sustainable software as SW that is easy to maintain, fulfils its intent over time, survives uncertainty, has the capability to endure, and supports relevant concerns (political, economic, social, technical, legal, environmental) such that it will continue to be available in the future, on new platforms, meeting new needs. The importance of software sustainability has been further underlined by a number of emerging initiatives and organisations such as the Software Sustainability Institute[153], Software Carpentry[154], funding initiatives from the National Science Foundation in the US[155], and the Engineering and Physical

---

[150] https://software.ac.uk/about

[151] https://danielskatzblog.wordpress.com/2016/09/13/defining-software-sustainability/

[152] http://wssspe.researchcomputing.org.uk/wssspe4/

[153] https://software.ac.uk/about

[154] https://software-carpentry.org/

[155] NSF A Vision and Strategy for Software for Science, Engineering, and Education. Available: https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12113

Sciences Research Council (EPSRC) in the UK[156,157]. In addition, a number of workshops[158,159,160,161,162,163,164] have emerged that are dedicated to exploring the topic of sustainable software and systems from a range of different perspectives.

Creating sustainable software can be challenging. Although there are many good practices associated with sustainability, there is no single set of accepted best practices. What is agreed is that SW needs to be maintained, otherwise it represents a risk for the project and the community at large. In the case of Science Gateways, it goes further: they need to be actively operated, maintained, monitored, patched, and secured, or they cannot guarantee support for the community they serve, and can become a cybersecurity risk. In either case, sustaining RS also requires HQP to do that as it doesn't just happen on its own.

There are a number of best practice guides available that provide recommendations for working through all stages RS lifecycle such as:

- The Netherlands eScience Center's comprehensive Software Development Guide[165] which includes a useful checklist for each stage of development.
- Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH)'s Software Quality Guide[166] that can be used by both software developers and software adopters to assess the quality and sustainability of software.
- The German Aerospace Centre's Software Engineering Guidelines[167] to support sustainable software engineering in research.

---

156 https://webarchive.nationalarchives.gov.uk/20200930204942/https://epsrc.ukri.org/funding/calls/csesoftwareforthefuture/

157 https://webarchive.nationalarchives.gov.uk/20200701162710/https://epsrc.ukri.org/funding/calls/softwareforthefuture/

158 https://sciencegateways.org/engage/focus-week/jumpstart#agenda

159 http://wssspe.researchcomputing.org.uk/welcome/

160 http://birgit.penzenstadler.de/re4susy/

161 https://www.knowledge-exchange.info/event/software-sustainability

162 https://www.software.ac.uk/attach/Research_Software_Sustainability_Report_on_KE_Workshop_Feb_2016_FINAL.pdf

163 https://software.ac.uk/programmes-and-events/collaborations-workshops

164 https://wosss.org/wosss21-home/

165 https://doi.org/10.5281/zenodo.4020565

166 https://github.com/CLARIAH/software-quality-guidelines

167 https://doi.org/10.5281/zenodo.1344612

- Productivity and Sustainability Improvement Planning (PSIP) Tools[168] that contains a collection of documents supporting an iterative planning process that helps improve scientific software developer productivity and product sustainability through improved software practices, processes, and tools.

4 RS Workforce

HQP in the context of RS refers to the professionals who are supporting and doing research, including researchers (including early career researchers, students and postdocs) and research support staff (e.g. sysadmins, research technical professionals (RTPs), software developers). The term HQP is regularly applied to two types of highly qualified personnel: new HQP via HQP development during the funded research, and HQP who are members of an existing research team or infrastructure organisation and already capable of conducting research activities and/or supporting research.

RS is developed through a highly distributed network involving various actors, including researchers, students, research software engineers (RSEs), librarians, university administrators, professionals in discipline-specific and non-profit organisations, and public servants in government. In Figure 3, Users & Research Communities who are the agents that are engaged in the activities with RS are omitted in all these activities. Above all, users (mostly researchers) are the essential actors who are engaged in the various activities throughout the research lifecycle. With respect to the actors, one can consider 3 broad groups of Users & Communities: 1). those that build Science Gateways (e.g. research software developers and teams) ; 2). those that create/deploy SW tools (e.g. sysops staff, SW developers) and support researchers in their use; 3). the researchers (e.g. non-professional/scientist using software) who perform research activities using RS by accessing either the gateways through their own computers or local resources and/or aforementioned SW tools. Given the relative “newness” of the RS community as a community of practice, these 3 communities are generally isolated one from another. The resulting community is diverse and lends itself to the development of HQP (e.g. junior researchers, trainees). This coalescence of tools and actors also makes it easier to support communities, create collaborative teams, and even surface synergies among multidisciplinary groups looking at grand challenges.

When we think of researchers who benefit from the potential that RS offers, we tend to imagine those working in STEM-related fields; however, RS is increasingly being used in the arts, humanities, and social sciences. As discussed in the “Future National Digital Research Infrastructure Landscape”[169], RS is an indispensable DRI component that is intertwined with the RDM and ARC components across all disciplines, driving a dramatic transformation in higher education, the scholarly landscape and research practice, and ensuring that Canadian

[168] https://betterscientificsoftware.github.io/PSIP-Tools/

[169] https://www.ic.gc.ca/eic/site/136.nsf/eng/home

researchers remain globally competitive and able to engage in international collaborations. Researchers and their associated research teams are not just using RS; they are also its prime producers. Some disciplines, such as biology, computer science and statistics, have spawned sub-disciplines, such as bioinformatics and data science, with their own journals, funding, and vibrant communities. In the humanities, those with a combination of research expertise and technical skills play a vital role in translating humanities inquiries into processes to which computers can contribute[170]. The development of RS comes largely from two groups: 1) highly trained SW developers who work with research groups and are typically employed as software engineers or developers and are increasingly referred to as RSEs; and 2) self-taught researchers who started as postdoctoral researchers, research staff/scientists or graduate students.

In addition, the Canadian University Council of Chief Information Officers[171] (CUCCIO) is another category of users and communities that are closely engaged with RS service delivery at the national level. CUCCIO is an incorporated federal non-profit, member-funded organisation consisting of Canada's provincial and regional associations of IT leaders in higher education. CUCCIO helps Canada's university community advance with the innovative and effective use of IT. CUCCIO's 62 member universities leverage opportunities to collaborate, advance and foster best practices, share information, explore new ideas, and build and sustain relationships with other groups to advance shared interests. As the only national conference for IT professionals in higher education, CUCCIO's annual conference CANHEIT (Canadian Higher Education Information Technology)[172] is hosted by one of its member universities, and has a focus on IT issues of interest to Canadian universities and colleges. As one of the associate members of the Alliance, there is an important opportunity for the Alliance to coordinate and collaborate with CUCCIO to advance the RS component.

5 RS in Canada

RS is developed and delivered through a highly distributed network involving various actors and stakeholders, including people (e.g. users and communities), organisations and associations, policy makers (e.g. funders and publishers) universities, private sectors and governments, etc. Given the number of players, the landscape can be challenging to navigate, but one useful approach is by jurisdiction: local, regional/provincial, national, and international, where each level presents some aspect of support or enablement, often in different and critical ways.

This chapter consists of two parts: Section 7.1 provides an overview of existing organisations and associations, post-secondary institutions, research hospitals and/or affiliated institutes, private

---

[170] Quinn Dombrowski, Tassie Gniady, David Kloster, Megan Meredith-Lobay, Jeffrey Tharsen, and Lee Zickel, Voices from the Server Room: Humanists in HPC, *Debates in the Digital Humanities,* forthcoming.

[171] https://www.cuccio.net/en/

[172] http://wps-canheit-prod.concordia.ca/

sectors and even journals that play a role in supporting or delivering RS services in Canada. Section 7.2 summaries the funders for supporting RS development and use.

5.1. RS within DRI and the Alliance

To align with the Alliance's goal of supporting a collaborative and agile digital research infrastructure community across Canada, the national RS activities within the Alliance, in parallel with the RDM and ARC components, are dedicated to enhancing researchers' capabilities so that they have access to appropriate research software that is robust, scalable, and sustainable, and embedded in an enhanced cybersecurity framework.

Modern data- and compute-intensive research is usually carried out using ARC infrastructure and systems, where huge amounts of data are stored and processed by RS. A significant proportion of RS is hosted in ARC, so the ARC and RS components can be considered to have a mutually dependent interaction. The development activities that are associated with debugging, data analysis and large-scale numerical modeling and visualisation are increasingly carried out directly on data also stored in that ARC environment. Access to ARC is transforming how we conduct research, deliver services, and provide advice and support, and ARC capacities are determined by the needs of RS processes and RDM. From another perspective, the design, development, and deployment of core ARC infrastructure itself is built on SW, and massively parallel and accelerator-enhanced ARC systems require very sophisticated and highly scalable RS products. The relationships between the various elements of RS, RDM and ARC are fundamentally interconnected and interdependent, and at all layers of the ecosystem. We cannot have one without the others, and we can no longer afford to have each of them functioning in their own siloed environments.

Cybersecurity can be considered as an integral component of the DRI ecosystem, encompassing mechanisms, strategies, policies, and practices that prevent unauthorized access to digital assets, and encompasses SW and hardware-based solutions for a wide range of wired and wireless communication networks. Cybersecurity is being advanced jointly by the Alliance and CANARIE in coordination with local institutions, and the global research and education network (GREN) by exporting innovations in platforms, software, and services to Canadian researchers.

5.2. Service Providers

For researchers to be able to find, access, reuse, and manage RS effectively, it is important to have a common framework of policies, procedures, protocols, standards, metadata, and DRI that are shared among a distributed landscape of individuals and organisations. This is especially true of a confederation like Canada, where substantial autonomy at the provincial level in areas of research and education means that coordination of efforts at the national level is critical, including national agreement on frameworks and practices. The Canadian research community has recognized this need, and two initiatives (see Section 7.1.1 and Section 7.1.2) have taken a leadership role in beginning to facilitate and coordinate Canada's RS-related activities at the national level.

5.2.1. CANARIE

Established in 1993, CANARIE (formerly the Canadian Network for the Advancement of Research, Industry and Education) is the not-for-profit corporation that coordinates Canada's National Research and Education Network (NREN), including the ultra-high-speed network connecting Canada's researchers, educators, and innovators to each other, and globally. CANARIE receives most of its funding from the Government of Canada. Driven by its mission to design, deliver and drive the adoption of digital infrastructure for research, education and innovation, CANARIE funds and promotes the development of research software tools; supports the development of policies, infrastructure and tools for research data management; collaborates to fund, implement, support and strengthen cybersecurity initiatives; provides access and identity management services to the research community; and assists Canada's startups and small/medium businesses with cloud resources and expertise in emerging technologies.

CANARIE's Research Software Program[173]

CANARIE's Research Software Program began in 2007 as a mechanism to fund development of research software tools that accelerate research discovery by enabling access to the digital infrastructure, and to allow researchers to focus on research instead of the supporting technology. Software provided with assistance from CANARIE's RS Program is designed to promote best practices, including those promoting code reusability to avoid duplication of SW development efforts. All software funded by CANARIE is accessible via the CANARIE Research Software Portal. The CANARIE Research Software Portal provides a central location for researchers and research software developers to discover software contributed by program participants, and by members of the greater research community. All software registered with the portal is free to use for research purposes by anyone, anywhere. CANARIE's Research Software Program invests in and supports both Research Software Platforms and Research Software Services in separate registries, with a goal of accelerating discovery by leveraging digital infrastructure resources. The Research Software Portal is open to all Canadian research software, not just CANARIE-funded projects, so any software developer is able to contribute platforms and services to the registry for use by other research teams. Since its inception, the RS Program has awarded $50.7M in funding to Canadian RS teams and has facilitated the development of Canada's RSE community.

CANARIE's RS program is critical in advancing and building capacity and the Canadian research software ecosystem. The success of some platforms (e.g. 3D Slicer[174], CANFAR[175]) that this program has funded is not only valuable to the individual communities but important to the RS ecosystem in Canada today. In addition to funding data-/compute-intensive communities, CANARIE is also crucial in funding non-traditional areas (e.g. CWRC[176] and Voyant[177] in digital

---

173 https://www.canarie.ca/software/

174 https://www.slicer.org/

175 https://www.canfar.net/en/

176 https://cwrc.ca/

177 https://voyant-tools.org/

humanities) where there used to be fewer sources of funding.[178,179,180,181] The impact that the Program has brought different domain researchers together to discuss RS practices is important: the focus on RS development, and building RSE culture and community, which is unlike other programs, is the key.

In order to promote reuse and best practices in research software development, since 2018 CANARIE has hosted an annual, discipline-agnostic event, the Research Software Workshop, more recently renamed the CRSC[182]. CRSC brings a community of RS professionals, developers and engineers from across the country together to exchange ideas and foster collaboration. CANARIE promotes the RS program to the community through periodic newsletters, blog posts,[183] and papers.

CANARIE is also an active participant in the international research software community. For example, CANARIE participated in International Research Software Engineering Leaders Workshop[184] and served as a founding member of international Research Software Alliance[185] (ReSA). In 2017[186] and 2018[187], in collaboration with the UK Software Sustainability Institute[188] (SSI), CANARIE conducted the annual Survey of Canadian Research Software Developers[189] to better understand the needs of the Canadian research software community. The survey results helped evolve and guide strategy for CANARIE's RS Program, ultimately ensuring that Canadian researchers have access to world-class software tools to accelerate their research. Preparation of the 2021 Canadian Research Software Survey is underway now.

In 2018, CANARIE launched a Local Research Software Support (LRSS) initiative[190,191]. Modelled partly on international efforts, funding is provided to local research software teams to develop, maintain, support, and evolve RS to equip Canadian researchers with SW tools and expertise

---

178 https://engagedri.ca/assets/documents/whitepapers/CSDH_NDRIO_WhitePaper.pdf

179 https://engagedri.ca/assets/documents/whitepapers/CADC_NDRIO_White_Paper.pdf

180 https://engagedri.ca/assets/documents/whitepapers/NDRIO_White_Paper_Medical_Imaging_Consortium.pdf

181 https://engagedri.ca/assets/documents/whitepapers/SBrown-Sustaining_DRI_in_the_Humanities.pdf

182 https://www.canarie.ca/event/crsc-2021/

183 https://www.canarie.ca/category/canarie-blog/

184 https://researchsoftware.org/

185 https://www.researchsoft.org/

186 https://github.com/canarie/developer_survey_2017

187 https://github.com/canarie/developer_survey_2018/blob/master/Results%20Graphs%20-%202018%20Canadian%20Research%20Software%20Developers%20Survey.pdf

188 https://www.software.ac.uk/

189 https://github.com/canarie/developer_survey_2018

190 https://www.canarie.ca/program-news/research-software/

191 https://www.canarie.ca/rdm-funding-call-recipients/

within the institution. The goal of the Local Research Software Support initiative is to promote the value of a strong RS team and encourage support at the institution for the long-term. It is also hoped that once such teams are in place in all regions, local software development efficiencies can be further scaled both nationally and internationally. McMaster is one of the three pilot projects funded by CANARIE's LRSS initiative, which was developed in 2018 and continued after a two-year pilot. McMaster's team is now self-funded as the Research & High Performance Computing (RHPC) initiative, which has evolved the pilot by funding RS teams to work directly with researchers. The other 2 pilot teams at University of Regina and at Carleton University have also committed to keeping the local teams in place post CANARIE funding. Additionally, all three teams added members (RSEs, domain-specific research support staff, etc.) that CANARIE did not fund during the pilot period, in response to high demand. With the success of this pilot, a second call was launched in 2020[192].

CANARIE'S Research Data Management Program [193]

CANARIE'S Research Data Management (RDM) Program (started in 2015) supports and facilitates the adoption of best practices for research data management[194] in compliance with international standards, including the provision of funding for SW development. CANARIE's RDM funding program has an RDM focus, although many of the same projects have received funding under both the RS and RDM programs, illustrating the overlap between the two. There have been two CANARIE RDM funding calls, with funding amounts of $3.2M and $2M in 2018 and 2019 respectively. RDM Call 1[195] focused on SW components and tools with strong support for the FAIR principles and the National Data Services Framework[196] (NDSF). A total of nine teams were funded.[197] Call 2[198] solicited proposals that would continue to focus on adherence to the FAIR principles and the NDSF, adding in support to enhance interoperability of data repositories and systems, nationally and globally. The call resulted in funding for 4 research teams.[199]

Like CANARIE's hosting of the CRSC, the RDM Program has hosted Research Data Management workshops in 2019[200] and 2020[201], providing an opportunity for the sharing and

---

192 https://www.canarie.ca/to-equip-canadian-researchers-with-powerful-software-tools-and-expertise-canarie-awards-3-6m-to-six-institutions-to-fund-local-research-software-development-teams/

193 https://www.canarie.ca/rdm/

194 https://www.force11.org/group/fairgroup/fairprinciples

195 https://www.canarie.ca/rdm/funding/funding-information-rdm-call-1/

196 https://doi.org/10.5281/zenodo.1035843

197 https://www.canarie.ca/rdm-funding-call-recipients/

198 RDM Call 2 is still underway: https://www.canarie.ca/rdm/funding/call2/

199 https://www.canarie.ca/canarie-awards-2m-to-research-teams-to-extend-the-interoperability-of-research-data/

200 https://www.canarie.ca/rdm/workshop2020/workshop2019/

201 https://www.canarie.ca/rdm/workshop2020/

exchange of ideas, and promoting opportunities to increase interoperability between infrastructures.

CANARIE will transfer both Research Software and Research Data Management programs to the Alliance by March 31, 2022.

5.2.2. Compute Canada Federation (CCF)

The Compute Canada Federation (CCF) is a partnership between Compute Canada[202], four regional consortia (ACENET[203], Calcul Québec[204], Compute Ontario[205] and WestGrid[206]) and institutions across Canada. The CCF delivers Canada's national ARC platform, which is designed to accelerate research and innovation by coordinating the deploying state-of-the-art ARC systems, digital infrastructure, storage, and software solutions. CCF's services and resources are freely accessible to Canadian academic research institutions, researchers, and their collaborators, and have played an important role in helping maintain and develop the country's leadership role on the international research scene.

CCF offers a catalogue of available research SW tools, libraries and environments[207,208]. In 2017, CCF adopted a common SW repository that was made available through their ARC clusters and thus provided a unified software environment to researchers. This SW stack can be deployed by users[209] on their own computers or clusters, allowing them to use multiple versions of the same software, and run their code simultaneously in multiple instances, exactly as they would on the CCF systems. In this way any CCF user can download the same software stack and use it as needed. This ensures better portability and reproducibility of scientific analyses. The CCF SW repository leverages a well-established technology developed by the CERN[210] physics community, CernVM-FS[211] (CVMFS), to deploy a distributed filesystem across the national clusters. Through the distributed filesystem, hundreds of SW tools and libraries can be installed and deployed across national sites in a reliable and efficient manner, the SW versions can be tracked centrally, and installation procedures can be automated. To access specific software, users simply invoke a command to load the appropriate module, which then loads the SW onto the computation nodes. With this distributed SW repository, the CCF is able to provide the

---

202 https://www.computecanada.ca/

203 http://www.acceleratediscovery.ca/

204 http://calculquebec.ca/en/

205 http://computeontario.ca/

206 https://www.westgrid.ca/

207 https://docs.computecanada.ca/wiki/Available_software

208 https://docs.computecanada.ca/wiki/Standard_software_environments

209 https://docs.computecanada.ca/wiki/Accessing_CVMFS#Installation

210 https://home.cern/

211 https://cernvm.cern.ch/fs/

Canadian academic research community with easy access to a large body of generic and domain-specific research software tools.

CCF is also working to recreate this environment in public clouds[212], taking advantage of a built-in-house open-source software project named Magic Castle. This framework takes cloud compute resources (e.g. AWS, Microsoft Azure, Google Cloud, OpenStack, and OVH[213]) and converts them into a virtual cluster or ARC infrastructure. After deployment, the user is provided with a complete ARC software environment that comes complete with the CCF SW repository, the Slurm[214] scheduling environment, a Globus Endpoint, JupyterHub, LDAP, DNS, and over 3,000 research software applications. Compute Canada staff are using this SW to deploy ephemeral clusters for training purposes, and it is also becoming popular with users for development, testing, and continuous integration.

In support of their operations, the CCF technical experts install, configure, and develop a large array of software systems; from a database (Compute Canada Database[215], referred to as CCDB) to track user accounts, infrastructure resources, computing allocations and run resource allocation competition, to sophisticated workload scheduler, monitoring and optimisation frameworks, advanced storage and cluster management systems, security tools and frameworks, to a helpdesk ticketing system for user support. The breadth of software expertise and systems found within the CCF is broad and constitutes a complex ecosystem at the service of Canadian academic research.

With more than 200 experts based at 37 partner institutions across the country, CCF provides direct support to Canadian researchers by offering consultation, expert support, and regional training[216] on a wide range of topics. Their national services[217] cover a wide range of areas from resource allocation to data movement (Globus[218]), and from visualisation to digital humanities. CCF also partners with researchers on projects involving an ARC component, providing in-kind resources (e.g. programmer time) over a prescribed term. For example, SHARCNET (one of the Compute Ontario consortia) has awarded dedicated programming support competition[219] to researchers in computational projects since 2008. In total, this popular program has supported 43 research projects from 10 different institutions over the past 13 years. Each project typically gets 4-6 months of support at up to 50% of a programmer's time. Other CCF partners offer similar programs, according to local demand and resource availability. In addition to the fact that any researcher can get access a default number of resources, those who work on bigger software

---

[212] https://github.com/ComputeCanada/magic_castle

[213] https://www.ovh.com/ca/en/

[214] https://slurm.schedmd.com/overview.html

[215] https://ccdb.computecanada.ca/security/login

[216] https://www.computecanada.ca/research-portal/technical-support/training/

[217]https://www.computecanada.ca/research-portal/national-services/

[218] https://docs.computecanada.ca/wiki/Globus

[219] https://www.sharcnet.ca/my/research/programming

platform projects can apply for even more computing and storage resources in the CCF's Research Platforms and Portals (RPP) Competition[220].

5.2.3. Institutional Hosting and Support

Post-secondary institutions, research hospitals and/or affiliated institutes across Canada have begun to invest in campus RS service delivery, by providing direct support to their researchers in terms of both infrastructure and access to RS tools, resources, research platforms, expertise, and training. The number of universities providing these services and the scale of investment required to support them is expected to increase significantly as new forms (e.g. collaboration, publication, and methods for visualising and analyzing data) of DRI emerge. There are a number of offices at higher education institutions that already have or will have responsibility for supporting these efforts at the local level, including:

- National labs located on university campuses (e.g. National Research Council Canada (NRC) labs[221]);
- Research offices (i.e. Vice-Presidents Research (VPR), research office staff, research policy committees);
- Information Technology Departments (ITD) or computing services departments (i.e. Chief Information Officers (CIO), academic computing personnel, institutional data governance committees);
- Libraries (i.e. subject/data librarians, metadata experts, curators/archivists, developers and support specialists, data management committees);
- External organisations (e.g. external service organisations "spun-off" from the institution, but dedicated to supporting the institution, such as Indoc Research[222] at the University of Toronto, Software Carpentry workshop at the University of Alberta[223])

Researchers themselves are often actively engaged in creating and delivering RS inside and outside their research programs, particularly in research labs or through domain-specific communities of practice and professional societies. For many institutions, specific technology-focused researchers, or heads of research institutes, can be heavily involved in local RS efforts.

However, every research institution has some level of local RS support, although the specifics are quite varied. Some institutions have a well-developed and extensive suite of services for RS, and in some cases these resources contribute to the provision of services at the regional and national levels. This model of a distributed support or service network can be seen in the Canadian DRI ecosystem and is reflected in the structure of some of the national players, such as CCF for

[220] https://www.computecanada.ca/research-portal/accessing-resources/resource-allocation-competitions/rpp/

[221] https://nrc.canada.ca/en

[222] https://www.indocresearch.org/

[223] http://computecanada.github.io/2016-04-28-ualberta/

ARC, and Portage in the RDM space. While there is not a similar distributed network in the RS landscape, this is emerging through the efforts of CANARIE, CCF, and some domain communities.

Many research institutions have IT units at different administrative levels that support RS in various ways. For example, Library IT teams sometimes provide basic support and training, and may also provide programming on a small scale. The Library may also host generalized software development services for what they see as common needs across the institution, and as a fundamental part of their support for open science. The central IT unit (or Technology Services) may also provide RS related services such as free or low-cost software purchase, installation and training, etc. Some departments, especially in STEM fields, have their own local IT support with dedicated staff who work extensively with research software. These teams can cater more to the specific needs of that research group.

Some universities also have dedicated centres for research computing that facilitate multidisciplinary discoveries through advanced computation, software engineering, data analysis, and other digital research methods. Their services may include advanced computing, data science, graphics and visualisation, software and systems engineering, cyberinfrastructure development, and training.

Research hospitals often have their own IT departments and local RS teams and develop their own research software that combines research with clinical practice. Research hospitals have unique and challenging issues, having to deal with data that involves patient information and therefore has strict privacy and security requirements. Research hospitals often have sizable research teams with large projects and due to the security and privacy constraints of their data they are often unable to use the DRI infrastructure provided by organisations like the Compute Canada Federation.

Colleges and institutes generally do SW development with industry partners, and sometimes with industry and university partners. Examples of college & university collaborations include the one-of-a-kind Master's Program in Ecological Restoration between British Columbia Institute of Technology (BCIT) and Simon Fraser University (SFU), George Brown College's partnership with the University of Toronto that gives students hands-on skills in manufacturing, the collaboration of Camosun College's Technology Access Centre (CTAC) and University of Victoria's Centre of Aerospace Research (CfAR) on the development of drone technology.

Industry-higher-ed collaborations in colleges include Entrepreneurship programs at BCIT, Nova Scotia Community College, Northern Alberta Institute of Technology, CEGEPs and Colleges in Quebec and Centennial College; and co-op experiences offered at Georgian College, etc.

Although fairly new in Canada, the idea of using external funding to facilitate the development of a local RS support teams is gaining traction: one example is CANARIE's Local Support Program[224] which provides multi-year funding for institutions to create local RS support teams. In addition, other similar services are emerging. For example, the Humanities Computing and Media

[224] https://www.canarie.ca/to-equip-canadian-researchers-with-powerful-software-tools-and-expertise-canarie-awards-3-6m-to-six-institutions-to-fund-local-research-software-development-teams/

Centre (HCMC)[225] at the University of Victoria has been around for many years as a permanent research and development office staffed by full-time software developers and experts in research and instructional design. Perimeter Institute not only delivers a world-class computing experience to its researchers, but also has a strong research software engineering practice and philosophy, which is unusual in the Canadian RSE community. As the Canadian and international RS ecosystem evolves, the opportunities to develop local RS expertise will evolve in tandem.

While this is an important evolution of RS services in the Canadian landscape, not all research institutions will have sufficient demand to warrant full time teams. National organisations like CANARIE that offer this support are important, helping to coordinate other RS software development models such as sharing RS teams across institutions.

In a similar context, most national research labs have their own local IT and RS support. While government labs are generally not eligible for federal funding, they are able to participate in collaborative research efforts. Despite this internal funding and support, national labs do look to the CCF and other national agencies for opportunities to share resources, typically by working with university-based collaborators. The Canadian Advanced Network for Astronomical Research (CANFAR) consortium[226] is one example, operated by the Canadian Astronomy Data Centre (CADC) at the Herzberg Astronomy & Astrophysics Research Centre, which is an NRC lab, while the PI is from Western University. The emergence of the Alliance presents Canada with an opportunity to achieve a closer and more interoperable strategy for the use of DRI across Canada's research enterprise.

There are a number of organisations and research groups that are dedicated to conducting Indigenous research[227] and contributing to educational websites[228]. Two platforms are Indigenous Knowledge Social Network[229] (SIKU), which is a mobile app and web platform by and for Inuit, and Indigenous Innovation Initiative[230], hosted at Grand Challenges Canada, which is an innovation platform aiming to remedy resource gaps, empower Indigenous Innovators and

---

[225] https://www.uvic.ca/humanities/hcmc/index.php

[226] https://www.canfar.net/en/

[227] e.g. Canadian Research Data Network: Indigenous Peoples (https://crdcn.org/indigenous-peoples), Centre for First Nations Governance (http://fngovernance.org/), CIER: Centre for Indigenous Environmental Resources (http://www.yourcier.org/), First Nations Information Governance Centre (http://fnigc.ca/), Institute on Governance (http://iog.ca/), iPortal: University of Saskatchewan (http://iportal.usask.ca/), Legacy of Hope Foundation (http://legacyofhope.ca/), OCAP (https://fnigc.ca/ocap-training/), Ontario Federation of Indigenous Friendship Centres (http://www.ofifc.org/), the Scow Institute (http://scow-archive.libraries.coop/), Statistics Canada: Aboriginal Peoples (https://www150.statcan.gc.ca/n1/en/subjects/aboriginal_peoples).

[228] including First Nation Profiles Interactive Map (https://geo.aadnc-aandc.gc.ca/cippn-fnpim/index-eng.html), iPortal: Indigenous Studies Portal (USask) (https://login.library.sheridanc.on.ca/login?url=http://iportal.usask.ca), Four Directions Teachings (http://fourdirectionsteachings.com/index.html), Library & Archives Canada: Aboriginal Heritage (http://www.bac-lac.gc.ca/eng/discover/aboriginal-heritage/Pages/introduction.aspx), Algonquian Dictionaries Project (www.atlas-ling.ca), Algonquian Linguistic Atlas (dictionaries.atlas-ling.ca), etc.

[229] https://siku.org/

[230] https://indigenousinnovate.org/

communities, and assist in capacity building of a new generation of indigenous innovation. Another recent development that is dedicated to supporting Indigenous research is the 2021 federal budget[231] that includes over $70 million for First Nations Data Governance Strategy development and implementation, as well as an investment of $8 million for Métis/Inuit baseline data capacity support and distinctions-based data strategies development. Additionally, SSHRC has a new commitment to funding research by and with Indigenous Peoples, including First Nations, Métis and Inuit peoples[232].

5.2.4. Private Sector

The private sector is a frequent supplier of RS to researchers. For example, research institutions are increasingly deploying data-/compute-intensive analysis via commercial cloud storage and computing, as well as services like IBM's Watson. Technology companies & startups may build RS to address their own business goals, and/or they may use tools developed in an academic environment, returning their own insights and enhanced tools in the process. Engaging the private sector in the conversation about the delivery of RS will be important, as it could be a key partner in the delivery of sustainable and cost-effective services. At the same time, universities and researchers will need to ensure that proper safeguards are in place to protect their ownership and primary stewardship roles for the RS that they generate (e.g. CLAs).

5.2.5. Journals

We also see some journals (e.g. Journal of Open Research Software[233] (JORS), Journal of open-source Software[234] (JOSS), Software Impacts[235], SoftwareX[236]) that are explicitly for RS. Journal publishers play a substantial role in directing and influencing researchers' approaches to RS, both through their official publishing policies[237], and their communities of practice. Elsevier, which publishes a large amount of scientific and medical research, has two journals dedicated to the publication of RS: SoftwareX for open-source RS and Software Impacts for peer-reviewed RS. Among the four journals, all three others[238,239,240] except SoftwareX are not transparent about the criteria they use to classify RS, nor the scope of RS they intend to include.

---

231 https://www.budget.gc.ca/2021/pdf/budget-2021-en.pdf

232 https://www.sshrc-crsh.gc.ca/society-societe/community-communite/indigenous_research-recherche_autochtone/index-eng.aspx

233 https://openresearchsoftware.metajnl.com/

234 http://joss.theoj.org

235 https://www.sciencedirect.com/journal/software-impacts

236 https://www.sciencedirect.com/journal/softwarex

237 https://www.chorusaccess.org/resources/software-citation-policies-index/

238 https://www.sciencedirect.com/journal/software-impacts/about/aims-and-scope

239 https://joss.readthedocs.io/en/latest/submitting.html#what-we-mean-by-research-software

240 https://openresearchsoftware.metajnl.com/about/#q1

5.3. Funders for RS Development and Use

Canadian research funding and awards[241] fall in two major categories: 1) scientific research funding; and 2) innovation and business research funding. The following list provides a representative set of organisations that support Canadian researchers by funding them explicitly or implicitly to develop and use RS. These organisations vary greatly in geographic scope (national, regionally and locally) and scope in terms of content (universal and disciplinary).

5.3.1. Nationally

5.3.1.1. CANARIE

Since its inception in 2007, the CANARIE's Research Software (RS) Program has championed the development of RS tools that accelerate discovery and simplify access to DRI by awarding a total of $50.7M in funding, with the typical ratio of applications to funding awards as 3:1. Among the 112 funded projects that cover many areas of academic research (180 disciplines), 49 projects involved international collaborators; 145 RS tools, services and platforms are available to researchers as a result of RS funding, with additional 617 demos/presentations/papers; and 1191 research teams benefited from funded software (related to 104 research teams provided with access to software per million dollars of funding). In particular, the RS tools that have emerged from this funding pipeline give Canada a first-mover advantage in leveraging DRI to accelerate discovery and spark innovation. CANARIE's RS program is critical not only in advancing and building capacity and the Canadian research software ecosystem but also in funding non-traditional areas where there used to be fewer sources of funding as well as in building the RSE culture and community in Canada.

CANARIE's RDM funding program, despite a RDM focus, demonstrates an overlap as some projects have received funding under both the RS and RDM programs. In terms of the total funding, CANARIE's RDM funding program has so far allocated[242] $4.5M to RDM platforms and an additional $1.8M to Portage. As of July 30, 2021, both CANARIE RDM funding calls have funded a total number of 13 teams; the funded projects cover 26 disciplines, made 12 tools, services and platforms available that benefit 628 research teams, and result in 62 demos/presentations/papers. Among them, three projects involved international collaborators.

5.3.1.2. Canada Foundation for Innovation (CFI)

Created by the Government of Canada in 1997, the Canada Foundation for Innovation (CFI) serves as an independent not-for-profit organisation that funds research tools and infrastructure that help Canadian researchers conduct cutting-edge research and technology development, as well as to build and sustain a full spectrum of research landscape in Canada. CFI funding is awarded ultimately to Canada's institutions (including universities, colleges, research hospitals and non-profit research organisations), not individual researchers. The CFI funds up to 40% of a project's research infrastructure costs, the remaining of which is then secured through investment from partners in the public, private and non-profit sectors. As of March 3, 2021, over $8.9 billion

---

[241] https://www.canada.ca/en/services/science/researchfunding.html

[242] "allocated" is used here because RDM Call 2 is still underway.

for 11,618 projects has been leveraged in research infrastructure in Canadian research institutions since its creation in 1997.[243]

There are six core funding streams that CFI provides to institutions, all of which allow RS development as part of the funding.

- John R. Evans Leaders Fund[244] (JELF): Supports institutions in securing the infrastructure resources (including key infrastructure and a portion of operating and maintenance costs) necessary to attract and retain top research talent. There are five different funding streams: the unaffiliated stream and four partnership (The Tri-agency Institutional Programs Secretariat's (TIPS) Canada Research Chairs[245] (CRC) and Canada Excellence Research Chair[246] (CERC) programs; The NSERC through its Alliance Grants[247] program; and SSHRC through its Insight Grants[248] and Partnership Grants[249] programs) streams.
- Major Science Initiatives Fund[250] (MSI): Supports unique, large-scale national research facilities with the ongoing operating and maintenance (O&M) that is necessary to produce excellent, world-class science and technology development. The purpose is to ensure that these facilities which are serving communities of researchers from across the country and internationally have robust governance and management policies and practices in place. The MSI program that is currently funding and sustaining CCF will be taken over by the Alliance on April 1, 2022.
- College-Industry Innovation Fund[251] (CIIF): Supports Canadian colleges and polytechnics in assisting emerging applied research infrastructure capacity and business innovation in fostering strategic partnerships with private sectors. There are two different funding

---

[243] https://www.innovation.ca/funded-projects

[244] https://www.innovation.ca/awards/john-r-evans-leaders-fund

[245] http://www.chairs-chaires.gc.ca/home-accueil-eng.aspx

[246] http://www.cerc.gc.ca/home-accueil-eng.aspx

[247] https://www.nserc-crsng.gc.ca/innovate-innover/alliance-alliance/index_eng.asp

[248] https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/insight_grants-subventions_savoir-eng.aspx

[249] https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/partnership_grants_stage1-subventions_partenariat_etape1-eng.aspx

[250] https://www.innovation.ca/awards/major-science-initiatives-fund

[251] https://www.innovation.ca/awards/college-industry-innovation-fund

streams: Stream 1 supports infrastructure requests with the goal of enhancing existing applied research and technology development capacity in colleges; Stream 2 is Innovation Enhancements program that is managed by NSERC and supports research infrastructure associated with an application for a five-year grant from the College and Community Innovation Program.

- Infrastructure Operating Fund[252] (IOF): Offers a 30% supplement to CFI capital awards that helps institutions cover a portion of the operating and maintenance costs of CFI-funded research infrastructure to ensure their optimal use.
- Innovation Fund[253]: Provides investment in leading-edge research infrastructure across all disciplines and areas of research.
- Exceptional Opportunities Fund[254]: Gives institutions the opportunity to seek funding for exceptional and time-sensitive infrastructure projects that would be missed within the normal CFI competition review process.

In addition, CFI's 2015 and 2017 Cyberinfrastructure initiative[255] aimed at designing and developing research data *infrastructures* via software development and involving multi-institutional consortia of institutions and their researchers, data scientists and software developers, and Compute Canada to conduct leading-edge computationally- and data-intensive research. This was CFI's first dedicated program in support of RS platforms, which are also funded under the Foundation's other programs, and it reflected the growing awareness of the challenges of sustaining infrastructure.

5.3.1.3. Other National Funding Agencies

Most of the national funding agencies are focused on funding research but not explicitly the development of research software or research infrastructure. That is why at the national/federal level, there are few funding sources targeted at developing research software as the primary outcome of a research project. Federal strategic initiatives like CANARIE and CFI provide funding for the development of RS tools, Gateways, and local and national RS support teams. Most national research funding agencies and initiatives don't typically fund projects to explicitly develop RS, but instead do so in an indirect manner by funding projects that investigate a domain-specific research problem and, through that research, develop RS as an outcome of the project. Examples

---

[252] https://www.innovation.ca/awards/infrastructure-operating-fund

[253] https://www.innovation.ca/awards/innovation-fund

[254] https://www.innovation.ca/awards/exceptional-opportunities-fund

[255] https://www.innovation.ca/awards/2015-and-2017-cyberinfrastructure-initiative

include Tri-agency (CIHR[256], NSERC[257] and SSHRC[258], Tri-Agency's Canada Research Continuity Emergency Fund[259] (CRCEF) program, Canada First Research Excellence Fund[260] (CFREF) initiative), New Frontiers in Research Fund[261] (NFRF), Canadian Space Agency[262] (CSA) (CSA's Flights and Fieldwork for the Advancement of Science and Technology[263] (FAST) funding, CSA's Space Technology Development Program[264] (STDP) funds, CSA's smartEarth[265] initiative, CSA's Innovative Solutions Canada[266] (ISC) program, CSA's international cooperation with the European Space Agency[267] (ESA) opportunity), Canada Research Chairs Program[268] (CRCP) and Canada Excellence Research Chairs[269] (CERC) program, Networks of Centres of Excellence[270] (NCE) initiative, MITACS' funding, Innovation Superclusters Initiative[271], and Canadian International Innovation Program[272] (CIIP). Genome Canada[273]/Quebec[274], while not focusing on RS funding calls like CANARIE, has supported

---

[256] http://www.cihr-irsc.gc.ca/e/193.html

[257] http://www.nserc-crsng.gc.ca/index_eng.asp

[258] http://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx

[259] https://www.sshrc-crsh.gc.ca/funding-financement/crcef-fucrc/index-eng.aspx

[260] https://www.cfref-apogee.gc.ca/home-accueil-eng.aspx#

[261] https://www.sshrc-crsh.gc.ca/funding-financement/nfrf-fnfr/index-eng.aspx

[262] https://www.asc-csa.gc.ca/eng/Default.asp

[263] https://www.asc-csa.gc.ca/eng/funding-programs/programs/fast/default.asp

[264] https://www.asc-csa.gc.ca/eng/funding-programs/programs/stdp/default.asp

[265] https://www.asc-csa.gc.ca/eng/funding-programs/programs/smartearth/default.asp

[266] https://www.asc-csa.gc.ca/eng/funding-programs/programs/isc/default.asp

[267] https://www.asc-csa.gc.ca/eng/funding-programs/funding-opportunities/default.asp?search=ESA

[268] https://www.chairs-chaires.gc.ca/home-accueil-eng.aspx

[269] https://www.cerc.gc.ca/home-accueil-eng.aspx

[270] https://www.nce-rce.gc.ca/Index_eng.asp

[271] https://www.ic.gc.ca/eic/site/093.nsf/eng/home

[272] https://www.tradecommissioner.gc.ca/funding-financement/ciip-pcii/index.aspx?lang=eng

[273] https://www.genomecanada.ca/en/

[274] https://www.genomequebec.com/

the development of many bioinformatics tools 275,276,277 and infrastructures278 to support genomics via funded research projects. The SSHRC Image, Text, Sound and Technology279 program was a dedicated funding stream for experimental tool development where some RS was developed, but the program has been discontinued and “tool” development is now subsumed under the Connection280 program which most researchers associate with conference or workshop support. As a result, there is less RS work happening in the Social Sciences and Humanities than there once was. In the context of Truth and Reconciliation, support for Indigenous research and data sovereignty in the 2021 federal budget has RS implications.

5.3.2. Regionally/Provincially

Traditional provincial/regional research funding agencies don’t explicitly fund RS: for example, Michael Smith Foundation for Health Research[281] (MSFHR) as a provincial health research funding agency aims to foster BC’s health research talent development and to optimize health research and knowledge translation in BC; Ontario Research Fund – Research Excellence (ORF-RE)[282], Early Researcher Awards[283], and the recent Ontario COVID-19 Rapid Research Fund[284] offer research and development funding for higher education to support research and innovation that helps cover the cost of research operations (e.g. salaries, supplies, etc.) alone. However provincial/regional funding agencies occasionally fund RS, such as Ontario Research Fund – Research Infrastructure[285]. But there are several provincial and regional organisations that have an impact on the development and delivery of RS services. Regional organisations may support RS development directly (the ACOA AIF[286,287] and related regional programs are one example), or indirectly via matching funding for other funding programs. There is also a recent trend to

---

275 https://www.genomecanada.ca/en/software-peptide-identification-and-quantification-large-mass-spectrometry-data-using-data

276 https://www.genomecanada.ca/en/rapid-prediction-antimicrobial-resistance-metagenomic-samples-data-models-and-methods

277 https://www.genomecanada.ca/en/target-big-data-informed-software-design-cis-regulatory-regions-controlling-human-gene-expression

278 https://www.genomecanada.ca/en/cancer-genome-collaboratory

279 https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/itst/research_grants-subventions_recherche-eng.aspx

280 https://www.sshrc-crsh.gc.ca/funding-financement/umbrella_programs-programme_cadre/connection-connexion-eng.aspx

281 https://www.msfhr.org/

282 https://www.ontario.ca/page/ontario-research-fund-research-excellence

283 https://www.ontario.ca/page/early-researcher-awards

284 https://www.ontario.ca/page/ontario-covid-19-rapid-research-fund

285 https://www.ontario.ca/page/ontario-research-fund-research-infrastructure

286 https://www.canada.ca/en/atlantic-canada-opportunities.html

287 https://www.canada.ca/en/atlantic-canada-opportunities/services/atlanticinnovationfund.html

reimagine largely health-focused provincial funding agencies to also support interdisciplinary and non-health focused research. The National Alliance of Research Health Funding Organizations (NAPHRO)[288] is an example of an organisation that is facilitating a collaborative approach across the country.

Provincial governments are also key players in supporting SW innovation in higher education institutions. This support varies from province to province to territory, but can include direct funding, matching funding, and a range of support programs. Examples include the Ontario Early Researcher Award, Innovation PEI[289], and Alberta Innovates[290]. Many of these provincial programs have regional economic development as a key investment strategy, often driven by institutional interests and capabilities. In the College sector, the private sector can be a substantial facilitator of SW innovation, with a substantial reliance on industry partners for funding and development focus.

5.3.3. Corporations & Foundations (Public, Private and International)

Corporate research partnerships are also a hub for innovation and entrepreneurship, providing extensive opportunities beyond the campus and bridging academia and industry, developing new technology[291] and sourcing talent. Many universities have created Technology Transfer offices to facilitate private sector partnerships, and spin-off activities from research, but most are focused on licensing and IP approaches. Open Source Program Offices (OSPOs)[292] have emerged internationally, but they have yet to see adoption in Canada. By providing commercialization assistance, research-driven ideas are advanced; commercial research investment is leveraged; and students get opportunities to explore careers and gain work experience while earning a university degree. More recently the private sector has started creating Open-Source Program Offices (OSPOs)[293], which are created to leverage the increasing value of open-source frameworks. While there are currently no OSPOs in Canadian higher-ed institutions, some are emerging elsewhere.[294] Some of these corporate research partnership engagements may also stem from the government funded or jointly funded programs, including some from NSERC, Mitacs[295], Ontario Centres of Excellence (OCE), SSHRC, etc.

Most foundations, charities and corporations (public, private or international), will usually fund research in a specific discipline but not in RS in particular. For example, the McLean Foundation[296]

---

[288] https://www.naphro.ca/

[289] https://www.princeedwardisland.ca/en/information/innovation-pei/information-and-communications-technology

[290] https://albertainnovates.ca/

[291] https://uwaterloo.ca/news/news/rogers-and-university-waterloo-partner-build-made-canada-5g

[292] https://www.linuxfoundation.org/resources/open-source-guides/creating-an-open-source-program/

[293] https://demand.fossa.com/what-is-an-ospo?source=Website

[294] https://blogs.library.jhu.edu/2020/12/hopkins-ospo-featured-on-impactful-open-source-podcast/

[295] https://www.mitacs.ca/en

[296] http://www.mcleanfoundation.ca/

makes grants in a wide range of areas, including social welfare, education, environmental conservation, health and the arts; the Graham Boeckh Foundation[297] (GBF) aims at transformational changes to the mental healthcare; the Jarislowsky Foundation[298] focuses on the promotion, support and advancement of excellence in education, medicine and the arts; Grand Challenges Canada[299] (GCC) supports integrated science and technology, social and business innovation in global health[300], humanitarian grand challenge[301] and assistance, and Indigenous Innovation Initiative[302]. Some programs fund research in cooperation with other agencies: Banting Research Foundation's Discovery Award Program[303] invests innovative health and biomedical research and data science, with additional co-funding in partnership with other organisations (e.g. Canadian Statistical Sciences Institute[304] (CANSSI), Dystonia Medical Research Foundation[305], Mitacs, University of Toronto McLaughlin Centre[306]). There is no doubt that RS plays an indispensable role in facilitating the overall funded research projects.

Some international funders and foundations have significant direct RS funding programs: The Chan Zuckerberg Initiative has an established Open Science funding program that is targeted at promoting the sharing of research outcomes including open-source research software[307]. Wellcome Trust[308] is another example, which not only invests in health research and global warming but has recently been funding RS directly through its Technology Development Grants[309]. The Trust also plays an important role in support of open access[310,311] and open data. In 2016 Wellcome Trust launched the Wellcome Open Research[312] open access publication platform and the Open Research Fund[313] that encourages making Wellcome-(co-)funded

---

297 https://grahamboeckhfoundation.org/en/

298 https://www.charitydata.ca/charity/the-jarislowsky-foundationla-fondation-jarislowsky/894744036RR0001/

299 https://www.grandchallenges.ca/

300 https://www.grandchallenges.ca/programs/global-health/

301 https://humanitariangrandchallenge.org/

302 https://www.grandchallenges.ca/programs/indigenous-innovation-initiative/

303 https://www.bantingresearchfoundation.ca/discovery-award-program/

304 https://canssiontario.utoronto.ca/

305 https://dystoniacanada.org/

306 http://www.mclaughlin.utoronto.ca/

307 https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/

308 https://wellcome.org/

309 https://wellcome.org/grant-funding/schemes/technology-development-grants

310 https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy

311 https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

312 https://wellcomeopenresearch.org/

313 https://wellcome.org/grant-funding/schemes/open-research-fund

research outputs FAIR. Specifically, the Open Research Fund call in 2021 focuses on incentivising sharing and openness in health research by covering development costs (building a new software tool or application) and licensing and computational (e.g. purchasing cloud-based storage) costs. For the purposes of research assessment, the Wellcome Trust is among the earlier adopters who utilize the DORA to improve the ways in which the outputs of scholarly research are evaluated[314].

For a more comprehensive list of funding sources for RS development and use, see Appendix H.

6 RS beyond Canada (Appendix I)

Appendix I presents a global environmental scan of the RS community landscape: initiatives, communities, companies, journals & publishers, non-profit entities, platforms, projects, registries/indexes/aggregators, groups, infrastructure, and services. This work is built on: 1) the output of the Research Software Alliance (ReSA) taskforce that identified and analysed the RS landscape[315,316]; and 2) the Scholarly Infrastructures for Research Software[317] (SIRS) by the EOSC Architecture WG that provided a broad overview of 9 representative European RS infrastructures. In addition to community activities, these organisations could be used to disseminate the latest news and updates, community calls[318], and funding opportunities[319], etc.

For Canadian researchers to be competitive in the research space, and to leverage RS created by their peers elsewhere, Canadian RS practices must coordinate with those globally, ideally with an element of coordination and collaboration from funders. Among the listed initiatives in Appendix I, there are several international organisations and/or events that Canadian RS organisations, associations and/or research communities have been actively engaged in at the international level. They include:

- UK Society of Research Software Engineering[320]
- US Research Software Engineer Association[321] (US-RSE)

---

[314] https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

[315] https://docs.google.com/spreadsheets/d/15JHqOxR4HIKHYe821IPvbxIuXP1zMjXKGEIJwB-GPqE/edit#gid=0

[316] https://doi.org/10.5281/zenodo.3699950

[317] https://www.softwareheritage.org/2020/10/23/towards-an-open-architecture-of-scholarly-infrastructures-for-research-software/

[318] https://us-rse.org/2021-02-19-community-call-education/

[319] https://www.fordfoundation.org/campaigns/critical-digital-infrastructure-research/

[320] https://society-rse.org/

[321] https://us-rse.org/, https://us-rse.org/usrse-map/

- Research Software Engineers[322]
- International Research Software Alliance[323] (ReSA)
- Research Data Alliance's (RDA) FAIR4RS working group[324]
- From an RS community perspective, CANARIE has largely played the role of representing Canadian RS at the international level.(e.g. CANARIE's roles in the Society of Research Software Engineering, international ReSA, and UK SSI; RDA FAIR4RS WG).

Despite minimal participation of Canada in global RS communities or organisations, more implementation details on reshaping of Canada's whole DRI to fit within an international context will emerge in many ways, such as part of the Researcher Needs Assessment Report, New National Service Delivery Model and Funding Model, or DRI Strategic Plan for 2022-2024.

7 Current Emerging Trends in DRI - RS Aspect

7.1. Cultural & Societal Perspective

An increasing number of research and civil society stakeholders advocate wider adoption of open science, which is discussed in Section 5. The promotion of open science ultimately requires a cultural shift in the way research is done and recognized. Effective open science depends on the creation of the openly available platforms, services, tools, and applications that enable researchers to mobilise knowledge. In that sense, specific elements of this shift include increasing collaboration and interaction between scientists and the development of technical infrastructure (e.g. Science Gateways or containerization) that promotes the development of emerging research practices in addition to conventional computing networks, as well as the development of "open-source science" practices.

7.1.1. FAIR Principles

RS is different from scholarly data and communications, but the high-level goal of the FAIR data Guiding Principles[325] also apply to RS. The FAIR principles are designed to ensure and improve optimal use, transparency, reproducibility and reusability of research data and other research objects. How the high-level FAIR principles should be applied to RS has emerged as a discussion topic over the last few years, and while it is generally accepted that the four foundational

---

[322] https://researchsoftware.org/2021/01/27/introducing-the-international-council-of-RSE-associations.html

[323] http://researchsoft.org/

[324] https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg

[325] https://www.force11.org/fairprinciples

principles[326] apply to RS, specific characteristics of software (e.g. executability, composite nature, continuous evolution, versioning) make it necessary to revise and extend the original guiding principles. The drivers, stakeholders, and incentives, whilst overlapping, are not identical for data and RS, and the variety of RS and its distribution channels poses a challenge when adapting the current FAIR data principles. Several groups are working towards the development of a set of FAIR Guiding Principles for RS, including the FAIR For Research Software Working Group[327] (FAIR4RS WG) which is co-led by RDA, FORCE11[328] and ReSA[329], and the efforts of the Software Sustainability Institute (SSI)[330]. The paper, *Towards FAIR Principles for Research Software*, published in 2020[331], is an example of work in this area. The recent report, Implementing FAIR for Research Software: Attitudes, Advantages and Challenges, published in 2021[332] discussed different topics around RS including attitudes, advantages, motivations, and barriers around the adoption of the FAIR principles for RS, current RS adoption practices across organisations, EDI, and software sustainability.

> *The first CANARIE RDM funding program had the adherence to the FAIR principles as a fundamental requirement.*

7.1.2. Training and Widening of the Use of Tools

Today most research has digital foundations, and the DRI requirements of fields previously not active in the traditionally STEM-focused ARC community are increasing dramatically. For example, specific challenges are posed by the broad array of methods that fall under the umbrella term of digital humanities (DH). DH brings together digital tools and methods with both large and modest data sets to advance inquiry across many disciplines, and frequently involves collaborative, trans- or interdisciplinary research that is closely connected to teaching and

---

[326] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016), 160018. doi:10.1038/sdata.2016.18.

[327] https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg

[328] https://www.force11.org/

[329] http://www.researchsoft.org/resa-taskforces-join-us/

[330] https://software.ac.uk/

[331] Lamprecht, A.L., Garcia, L., Kuzak, M., Martinez, C., Arcila, R., Martin Del Pico, E., Dominguez Del Angel, V., van de Sandt, S., Ison, J., Martinez, P.A. and McQuilton, P., 2020. Towards FAIR principles for research software. *Data Science*, *3*(1), pp.37-59. https://content.iospress.com/articles/data-science/ds190026

[332] https://ssi-cw.figshare.com/articles/conference_contribution/Implementing_FAIR_for_research_software_attitudes_advantages_and_challenges/14453031

dissemination (e.g. of datasets, remediated historical sources, thematic research collections, and other digital scholarship). DH researchers adapt both tools and algorithms from other fields, for instance AI, ML, and gene sequencing for activities such as text mining, image processing, and data analysis, as well as developing tools themselves. The humanities and social sciences encompass numerous disciplines with multiple methods of inquiry, types of data, and diverse research software needs, so there is no one-size-fits-all solution for these underserved fields when it comes to providing equitable access to ARC; some work may not look like ARC from a traditional perspective. Because computational methods and data analytics are yet to be integrated into most humanities curricula, there is a pressing need for other forms of training in appropriate methods. When it comes to RS and training in its use, existing and emergent community-led initiatives (e.g. Digital Humanities Summer Institute[333] (DHSI) and other Canadian training initiatives, Alan Turing Institute's Digital Humanities & Research Software Engineering summer school[334], CODATA-RDA School of Research Data Science[335]) and communities of practice can serve as partners in training with the Alliance to meet researcher community needs.

7.1.3. Software and Algorithmic Bias

As algorithms harness a variety of macro- and micro-data collections to build ML and AI tools, more attention is being paid to the study of algorithmic bias[336]. Bias can emerge due to many factors, including the design of the algorithm itself, the unintended use of incomplete or unrepresentative training data, or the way data are coded, collected, selected, or used by an algorithm. Algorithmic bias can be found across the whole spectrum of tools, including most AI-based platforms (e.g. search engine results, facial recognition, social media platforms, criminal justice algorithms), and can have impacts ranging from inadvertent privacy violations to reinforcing social biases (e.g. race, gender, sexuality, and ethnicity) and raising ethical concerns.[337] The study of algorithmic bias can be partially addressed in legal frameworks, such as the European Union's General Data Protection Regulation[338] (GDPR), but more comprehensive regulation, new pedagogical approaches to teach responsible data science[339], and bias detection strategies are needed as emerging technologies become increasingly advanced and opaque. The RS community can clearly benefit from the use of ML and AI SW, but the approach to integration of the algorithms and data needs special understanding and attention.

---

[333] https://dhsi.org/

[334] https://www.eventsforce.net/turingevents/frontend/reg/thome.csp?pageID=23222&eventID=72&traceRedir=2

[335] https://codata.org/initiatives/strategic-programme/research-data-science-summer-schools/

[336] https://en.wikipedia.org/wiki/Algorithmic_bias

[337] https://www.jstor.org/stable/j.ctt1pwt9w5

[338] https://gdpr.eu/

[339] https://link.springer.com/article/10.1007/s40593-021-00241-7

7.1.4. Citizen Science

While open science opens the doors of academia to the world, citizen science[340] invites the world in. Within the framework of open science, the interaction between citizens and scientists reduces the gap between the two, informs the traditionally insular approach to research, and can improve the profile of science in society. In this way, experts and non-experts solve problems of common interest together, according to the rigor of the scientific method. RS is facilitating citizen science via efforts that enable volunteers to participate in data collection activities, facilitate data analysis and, in the case of open-source software, co-develop the tools that underpin research processes.

7.1.5. RSE as a Professional Role

The concept of a research software engineer (RSE) as a professional role was developed only in the last decade. The birth of the term RSEs can be attributed to the RSE movement in the UK, combined with a series of SSI-hosted workshops that identified that software developers lacked more than just recognition. RSE is an actively growing industry with an increasing number of vacancies and roles. France, for instance, has codified categories of academic research support occupations (e.g. “research engineer”[341] positions associated with handling data across all disciplines[342]). Although the world is using the term “RSE”, unfortunately Canada doesn't legally allow the use of the term “engineers” but uses the term “research Software developers” instead. The importance of high quality RSEs is becoming more and more prominent, both within institutions and funding agencies (e.g. the EPSRC Software for the Future[343], NSF CSSI[344] (previously NSF SI^2 funding[345])), thanks to initiatives like the CANARIE programs, SSI, Society of Research Software Engineering, and Hidden REF[346], but also in the general public given the role of RSE in marshalling rapid responses to the COVID-19 pandemic[347].

7.2. Technology Evolution

In addition to the cultural and societal perspectives mentioned in Section 8.1, a number of newer and evolving technologies and digital trends are gaining momentum and climbing in popularity and are impacting and being impacted by RS.

---

340 https://en.wikipedia.org/wiki/Citizen_science

341 https://www.enseignementsup-recherche.gouv.fr/cid23194/ingenieur-de-recherche.html

342 https://metiersit.dsi.cnrs.fr/index.php?page=cartofamille&codeBAP=D&codeFamille=A

343 https://epsrc.ukri.org/research/ourportfolio/themes/researchinfrastructure/subthemes/einfrastructure/software/

344 https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505505

345 https://www.nsf.gov/pubs/2013/nsf13511/nsf13511.htm

346 https://hidden-ref.org/

347 https://society-rse.org/rse-contributions-to-covid-19-research/

7.2.1. ML and AI

Nowadays ML and AI have become one of the most intriguing areas of computer science research and the technology (e.g. machine control, machine learning, deep learning (DL), and neural networks) is evolving quickly, particularly over the last five years as computing hardware and machines have evolved in step with the SW. From the RDM perspective, ML, AI, and other modern statistical methods are providing new opportunities to operationalise previously untapped and rapidly growing sources of data. AI and ML software solutions are not just a 'trend' anymore, but a dominant force in fields as diverse as healthcare, finance, education, transportation, as well as in RS development, such as automated testing, and automated code generation. There's no shortage of opportunities to develop real-world applications of the technology, and there's immense scope for break-through moments in this field. In 2017, the Government of Canada appointed CIFAR to develop and lead a $125 million Pan-Canadian Artificial Intelligence Strategy, the world's first national AI strategy.[348] CIFAR works in close collaboration with Canada's three national AI Institutes (Amii[349] in Edmonton, Mila[350] in Montreal, and the Vector Institute[351] in Toronto), as well as universities, hospitals, and organisations across the country. Interdisciplinary groups pursuing research and impact projects involving ML and AI would also benefit from explicitly addressing a series of questions concerning transparency, reproducibility, fairness, accountability, ethics, and effectiveness (TREE[352], FAT[353]), as well as the previously mentioned FAIR Principles.

In responding to support Pan-Canadian AI Strategy's National Program of Activities to support cross-collaboration among Canada's top AI scientists and inspire high-risk, high-reward areas of research: *Nine projects were funded through the CIFAR AI Catalyst Grants program.*

7.2.2. Cloud Computing

As an on-demand over-the-Internet approach to accessing computing resources (e.g. applications, physical/virtual servers, data storage, databases, software, analytics, development tools, infrastructure, networking capabilities, intelligence), cloud computing has become a dominant force in resource delivery systems. Organisations are increasingly implementing cloud services for their researchers (e.g. Compute Canada Cloud[354]), leveraging the advantages of different costing models, high performance and productivity, reliability, manageability, elastic

---

[348] https://cifar.ca/ai/

[349] https://www.amii.ca/

[350] https://mila.quebec/en/

[351] https://vectorinstitute.ai/

[352] https://www.bmj.com/content/368/bmj.l6927

[353] https://www.fatml.org/

[354] https://www.computecanada.ca/research-portal/national-services/compute-canada-cloud/

scale, data centralization, proper security, etc. Most cloud computing services fall into four broad categories: infrastructure as a service (IaaS), platform as a service (PaaS), serverless (also referred to as function-as-a-service), and software as a service (SaaS), also sometimes referred to as the cloud computing stack because they can build on one another. Although cloud computing has been around for a while, its usage in terms of research activities and from RS perspective is still emerging and not broadly adopted. The themes related to RS in the field of cloud computing include business (e.g. cloud computing adoption, privacy, legal and ethical issues, trust, cloud cryptography and security for protection against foreign attacks and breaches), technology (e.g. cloud analytics, cloud computing platforms, service and architecture, software development, cloud load balancing, containerization, edge computing, data centre management, security), and cloud computing application domains (e.g. in education, e-science, health, knowledge management, open-source software, mobile cloud computing, and green cloud computing focusing on making virtual data centres and servers).

Cloud computing has particular relevance to the Digital Humanities for reasons ranging from the frequently iterative and collaborative nature of research activities to the interactivity of both tools and data such as games, and the requirements of some kinds of data, like linked open data, to be accessible 24/7. Netlytic[355] provides online tools for the collection and analysis of social media. The Archives Unleashed project[356], including the Archives Unleashed Cloud, addresses the need to make large volumes of defunct Web data usable by researchers in HSS regardless of whether they can program. It should also be noted that although they might not meet conventional definitions of RS, community building and research dissemination sites play important roles in the HSS infrastructure ecosystem in circulating grey literature and fostering dialogue and exchange. These may be domain-focused, such as the Network in Canadian History & Environment[357] (NiCHE), or more broadly targeted, as in the Canadian HSS Commons, now in beta, which adds project-management functionality into its FAIR-principles based publications and networking platform. The Digital Humanities are therefore heavier users of Cloud as opposed to batch computing; the most-used platform on the Compute Canada Cloud is the Voyant Tools suite for text analysis and visualisation.

7.2.3. Infrastructure as Code

Maintaining cloud infrastructure can be a challenging task, as complicated setups can become fragile at scale without the right architecture. Despite that complexity, there are many benefits offered by cloud computing and virtualization technology, and one of the most important may be improvements in managing infrastructure in a consistent and replicable way. One way of doing this is through Infrastructure as Code[358] (IaC), which creates an abstraction layer for packaging, organising, and building out cloud environments on demand through automated scripts, based on an existing model rather than having to rebuild from scratch each time. For those working with complex cloud builds, IaC makes for simpler management in the long run. As a cloud-enabled technology, IaC also makes it possible to easily scale and redistribute infrastructure on demand.

[355] https://netlytic.org/

[356] https://archivesunleashed.org/

[357] http://niche-canada.org/

[358] https://en.wikipedia.org/wiki/Infrastructure_as_code

This is critical to research reproducibility as it allows one to define and model infrastructure in a formal way through software; store the infrastructure code in version control systems; write configuration code as documentation; reuse parts of the architecture in future projects or in any desired cloud; and facilitate sharing. From an automation standpoint, IaC can help ensure efficiency in the deployment of resources while offering quick and consistent configuration and flexibility, while lowering the overall cost of infrastructure management. Resources can be distributed via the cloud or on-premises as needed, which can come in handy during major crises — such as the start of the COVID-19 pandemic, when some organisations or businesses found themselves suddenly needing to scale up.

7.2.4. Internet of Things (IoT)

The IoT is an emerging paradigm, using smart devices in an internet connected context to provide innovative solutions to various challenges and issues related to business, governmental and public/private industries across the world. From wearables to connected kitchen appliances, IoT devices are everywhere. The global IoT market is expected to exceed $4T USD by 2025[359].

IoT can provide benefits to a broad range of disciplines such as finance, environment, industry, education, transportation, but especially in health sciences (also referred as smart healthcare or eHealth). IoT driven eHealth (e.g. remote health monitoring, emergency notification, digital imaging, mobile apps for editing/displaying health records, robotic surgery) can help create a digital healthcare system (e.g. infrastructures and platforms) to collect data and connect available healthcare resources and services. The potential with IoT can be seen in other domains, such as smart traffic management, security and surveillance, agriculture and smart farms, smart manufacturing, smart devices (e.g. wearables, vehicles, robots), smart cities & homes, retail, energy & grid control, etc. IoT, combined with other innovations such as data analytics and AI/ML/DL, is important to society.

Although there is a wealth of crucial research studies into the IoT, there are still many challenges and issues that need to be addressed to achieve its full potential. One of the most important is the security and privacy of IoT devices due to threats, cyber attacks, risks, and vulnerabilities. There are also issues with interoperability, due to the rapid proliferation and heterogeneous nature of different IoT technologies. One can consider four interoperability levels with respect to the IoT: technical, semantic, syntactic, and organisational[360]. Other issues with IoT include scalability, availability, reliability (e.g. computing, device, communication, network, application reliability Quality of Service (QoS)).

IoT also creates programming challenges. Programming across the IoT landscape is very challenging, having to accommodate multiple devices, each being different in terms of standards, APIs, and metadata formats. In addition, programming for wearable devices is quite different than programming for servers and desktops, requiring low-level languages (e.g. C) and specific software skill sets (real time programming).

---

359 https://www.mckinsey.com/featured-insights/internet-of-things/how-we-help-clients#

360 Van-der-Veer H, Wiles A. Achieving technical, interoperability-the ETSI approach, ETSI White Paper No. 3. 2008. http://www.etsi.org/images/files/ETSIWhitePapers/IOP%20whitepaper%20Edition%203%20final.pdf

7.2.5. Edge Computing

In our data-heavy future, with billions of devices connected to the internet, faster and more reliable data processing will become crucial. Edge Computing (EC) is also a new and emerging field: a distributed IT system with decentralized processing power, which optimizes the resource usage in cloud computing with the idea that the data are processed nearer to the source - at the edge of the network - rather than on a centrally managed platform or the data warehouses. Moreover, it saves bandwidth and improves the security of the data due to its proximity to the source and elimination of late response time and latency without affecting productivity. According to Grand View Research, the global EC market was worth $3.5 billion USD in 2019 — and that number is expected to grow to $43.4 billion USD by 2027[361].

EC complements cloud computing in a hybrid IT environment. While cloud computing leverages centralized data centres, EC leverages distributed micro data centres at the edge of the network, where data are used closer to the source. EC is necessary to address shortcomings in cloud-based applications and services with respect to performance and regulatory requirements. It's an issue because the trend toward digitization to improve efficiency and performance is fueling demand for applications that require peak performance, particularly IoT applications. IoT applications often require substantial bandwidth, low latency, and reliable performance, while meeting regulatory and compliance mandates, making them classic candidates for EC. The principle behind the applications of IoT and EC is the same, regardless of the exact implementation: devices or sensors at one end sending data to an edge data centre for processing and perhaps some analytics, then to a more centralized application (often in the cloud) that delivers the service. Software solutions for virtualization, such as Virtual Machine (VM), containers, and their management or migration tools, have been active enablers of EC systems. More recently, software-managed networks also allow further exploration of the potential of EC by facilitating appropriate technologies (e,g., Software-Defined Networking (SDN), Network Function Virtualization (NFV), and Overlay network).

Despite continued research and improvement, EC technology is still relatively young, and challenges around device capabilities — including the ability to develop software and hardware that can handle computational offloading from the cloud — are likely to arise. Enabling machines to toggle between a computation that can be performed at the edge and one that requires the cloud is also a challenge. The upcoming commercialization of the fifth generation mobile communication network (5G) provides new opportunities for the development of EC, particularly in real-time environments. Promising challenges for future research and technologies in EC include mobility management, data security and privacy protection, heterogeneity, reliability, storage data models, and simulation environments. While some use cases may prove the value of edge computing more clearly than others, the potential impact on the rising and connected ecosystem could be substantial.

7.2.6. Quantum Computing

The power of quantum research harnesses the quantum laws of nature to develop powerful new technologies. Interdisciplinary quantum research spans theory and experiment; fosters collaborations across science borders; and focuses on core research areas such as quantum

---

[361] https://www.grandviewresearch.com/industry-analysis/edge-computing-market

computing[362], quantum communication, quantum sensing and quantum materials. Quantum computing (QC), by harnessing the quantum behaviour of atoms, molecules and nanoelectronic circuits for a radically different, and fundamentally more powerful way of computing, could spur the development of new breakthroughs in science, materials, and medicine. Potential QC applications and impacts include security (e.g. quantum cryptography), optimisation (e.g. quantum annealing[363] and adiabatic quantum optimisation [364], unstructured search problems), quantum supremacy[365], simulation of physical, chemical and biological processes, and drug and material design. Although this technology is truly emerging, it is still in its infancy with many unknowns. For example, how to write traditional codes and have parts that actually leverage quantum computing and do it in a meaningful and effective way is a huge challenge as there are not many frameworks that allow easy programming in that area. Plus, most researchers would not know how to approach a quantum computer or what to do with it. So, the impacts may require a whole new method of programming and whole new languages.

7.2.7. Data Analytics

Since the 1990s, data analytics has facilitated a revolution in the production of knowledge within and beyond science, by enabling novel, highly efficient ways to plan, conduct, disseminate and assess research, as well as in the creation of novel ways to produce, store, and analyse data. Data analytics has facilitated the emergence of the field of data science, which brings together computational, coding, algorithmic, statistical, and mathematical techniques plus domain expertise to extrapolate knowledge from data analytics. From finance to healthcare, data analytics are pervasive, and along with the push for open science and open data, have encouraged the sharing and interlinking of heterogeneous research artifacts using DRI. The availability of vast amounts of data in machine-actionable formats provides an incentive to create efficient procedures to collect, organise, visualise and model these data. These data-focused infrastructures, in turn, serve as platforms for the development of AI approaches, with an eye to increasing the reliability, speed and transparency of processes of knowledge creation. Researchers across all disciplines can benefit from the ability to link data from diverse sources, improving the accuracy and predictive power of scientific findings, and helping to identify future directions of inquiry. As exemplified by the rise of dedicated funding, training programs and publication venues, data analytics is widely viewed as ushering in a new way of performing research and challenging existing understandings of what counts as scientific knowledge.

The adoption of data analytics technologies (e.g. Apache Hadoop, Spark, R, Python, data lakes, NoSQL databases, predictive analytics, streaming analytics, EC, blockchain, in-memory computing, data security, AI, data governance, prescriptive analytics) is unlikely to slow anytime soon. For example, large-scale simulation is one of the big beneficiaries of big-data analysis and unprecedented computing power. Machine learning is being merged with analytics and voice responses, while AI drives deeper insights and increasingly sophisticated automation. open-source applications have come to dominate the data analytics space, and that trend looks likely

---

[362] https://en.wikipedia.org/wiki/Quantum_computing

[363] https://en.wikipedia.org/wiki/Quantum_annealing

[364] https://en.wikipedia.org/wiki/Adiabatic_quantum_computation

[365] https://en.wikipedia.org/wiki/Quantum_supremacy

to continue. Data analytics solutions can make a decisive contribution to the debate on climate change and approximate events from the Big Bang to the present time of the universe.

7.2.8. Cybersecurity

Today, all large-scale research infrastructures are dependent on information and communication technology (ICT) resources, providing new possibilities for geographically distributed collaboration and sharing. This ICT dependence has increased the need to find synergies, and to develop ways to tackle the ICT challenges at a foundational level, providing effective and cost-efficient services that can be of wide and general use. The DRI for networking, ARC, RDM, and cloud computing are evolving rapidly, and consolidation and integration is facilitating service delivery to internationally distributed research projects. Beyond e-infrastructures and digital devices for network, computation and RDM, it is necessary to develop specific research infrastructures in the domain of computer sciences, supporting the experimentation of disruptive systems including e-infrastructures, hard-/middle-/software, protocols, computing and cybersecurity issues. RS developers haven't typically considered the importance of cybersecurity in their software or e-Infrastructure design. But because of the increasing number of intrusions and breaches of confidential information and health data in particular, research software engineers almost have to move this to the top of the priority queue and make sure they have robust mechanisms that institute good security. In today's hyper-connected digital research world, cybersecurity is no longer optional, making this another growing strand of computer science research. This has been done as a co-design effort between all stakeholders where new needs of the researchers lead the way to innovation efforts.

7.2.9. Research Reproducibility

It is critical, especially in a crisis such as the current pandemic, for RS to produce results that can be reproduced. The holy grail of full SW reproducibility has been the object of much attention in many disciplines, with tools developed ranging from virtualisation or containerization environments, to notebooks. This could include automatic logging of all parameter values (including setting random seeds to predetermined values), as well as establishing the requirements in the environment (e.g. dependencies). Examples of systems currently being used by researchers to capture and replicate research environments include virtual machines and containers, cloud computing (7.2.2), and infrastructure as code (7.2.3). These systems now become part of the new delivery of sustainability (preservation) model. In addition to the systems used in the delivery model described above, RS pipelines and scientific/computational workflows[366] are being built for automating the management of tasks in a complex analysis, enabling computational reproducibility. For more examples, see Appendix F.

RS also supports and assists in RDM (e.g. data dissemination and sharing), which is a crucial part of reproducibility. The growing adoption of programs like Jupyter and R markdown, which produce dynamic computing documents containing live code and descriptive text, are also allowing data reusers to interact with open data directly. Support for these SW tools may

---

[366] e.g. Kurator (https://github.com/kurator-org), Nextflow (https://www.nextflow.io/), nf-core (https://nf-co.re/), Vistrails (https://www.vistrails.org/index.php/Main_Page), Galaxy (https://galaxyproject.org/)

increasingly be provided by repositories to support computational reproducibility (e.g. the Git - Zenodo - Binder integration[367]).

7.2.10. Exascale Computing and RS

Exascale computing refers to the ability of a single machine to perform 10^{18} floating point operations per second (an exaflop). Such architectures are typically hybrid architectures that consist of a hierarchy of parallel components (nodes, sockets, cores) with sockets often consisting of heterogeneous components that include highly parallel co-processors such as GPUs[368]. Each of these components typically has on board memory, resulting in complex memory hierarchies with different memory access performance levels. These components are then connected by interconnects with varying bandwidth and communication latency.

Such architectures are extremely difficult to program efficiently, with traditional parallel programming techniques and libraries (OpenMP, MPI, OpenCL, CUDA etc) optimized for a single level of this hierarchical architecture (e.g. MPI for parallelism across nodes/cores and CUDA for parallelism across multi-core GPUs). Writing RS for exascale class systems requires a unique skill set that not only involves knowledge on using parallel libraries such as MPI and CUDA, but knowledge and skills on combining such libraries and optimizing codes such that the application utilizes the interconnects between the components in the hierarchy. Achieving high efficiency on exascale class systems requires the avoidance of bottlenecks at all levels of this heterogeneous and hierarchical architecture. Although the research community is working on new programming models for these complex heterogeneous architectures, no such solution is currently widely accepted and used. Exascale computing is discussed in more detail in the Alliance ARC position paper.

8 Current State Assessment

8.1. Strengths

8.1.1. HQP

In Canada, we have a combination of good support for research programs that include software development efforts, and strong computer science curricula nationally. That includes university curricula in software engineering[369], institutional SW research centres and institutes[370], other

---

[367] https://blog.jupyter.org/binder-with-zenodo-af68ed6648a6

[368] http://www.intertwine-project.eu/introduction-interoperability

[369] e.g. Ontario Tech University (https://ontariotechu.ca/index.php), UVic's Software Engineering Program (https://ontariotechu.ca/index.php), Master of Engineering Leadership (MEL) in Dependable Software Systems at the University of British Columbia (https://apscpp.ubc.ca/programs/mel/dependable-software-systems/)

[370] e.g. Concordia University's Software Engineering Research Centre (https://www.concordia.ca/research/software-engineering-research-centre-serc.html), Canadian Statistical Sciences Institute (CANSSI) (http://www.canssi.ca/), PIMS (https://www.pims.math.ca/),

research groups[371], co-op programs[372], and industry-institution collaborations in colleges[373]. There is also a strong foundation for RS development and RSEs via CANARIE's activities, including the RS Software Conference[374], participation in the International Research Software Engineering Leaders Workshop[375], acting as a founding member of Research Software Alliance[376] (ReSA), and launching the Local Research Software Support initiative. As an example of a RS-focused initiative, the Perimeter Institute not only delivers a world-class computing experience to its researchers, but also has a strong research software engineering practice and philosophy: unusual in the Canadian RSE community. As other academic research groups further develop and enhance their RSE teams, there will be an opportunity to collect these teams into a national RSE organisation.

There is a large SW development talent pool, especially within industry, which attracts more qualified people to a thriving and vibrant ecosystem. For example, Mitacs has been in partnerships with 70 Canadian universities, 6,000 companies, and both federal and provincial governments, to operate research and training programs in support of industrial and social innovation. Jointly funded by the federal and provincial governments, as well as academic and innovation partners[377], Mitacs supports more than 20,000 research projects, over 3,600 international research collaborations, and has trained more than 33,000 student and postdoc participants over the past 20 years. A focus for Mitacs is bridging the skills gap between industry and academia, facilitating the development of a strong talent pool. For many years NSERC offered Engage grants, which also facilitated industry-university partnerships, including many projects which saw developers in universities create code for companies based on their specific requirements. Engage grants are no longer available for universities but are still available for colleges and through SSHRC.

A deep talent pool and vibrant software industry can be both a strength and a challenge, since research has to compete with industry for HQP, who are typically better paid in industry. Despite that competitive disadvantage, universities and colleges are able to provide other benefits (e.g. quality of life, work that aims to have a social rather than financial impact) that can attract good developers who tend to stay around for the long-term.

Fields Institute for Research in Mathematical Sciences (http://www.fields.utoronto.ca/), Centre de recherches mathématiques (CRM) (http://www.crm.umontreal.ca/en/index.shtml), Atlantic Association for Research in the Mathematical Sciences (AARMS) (https://aarms.math.ca/)

[371] e.g. Indoc (https://indocresearch.org/aboutus.html), High Energy Physics group at SFU (http://hep.phys.sfu.ca/), Robertson Library's software development team at UPEI (https://library.upei.ca/vre), Research Commons at UBC(https://researchcommons.library.ubc.ca/)

[372] e.g. University of Waterloo (https://uwaterloo.ca/co-operative-education/), UVic's co-op education program (https://www.uvic.ca/engineering/software/co-op/index.php)

[373] e.g. Mohawk (https://www.mohawkcollege.ca/)

[374] https://www.canarie.ca/software/canadian-research-software-conference/

[375] https://researchsoftware.org/

[376] https://www.researchsoft.org/

[377] https://www.mitacs.ca/en/about/partners

8.1.2. National Funding and Partnerships

In Canada, there are two RS programs at the national level: CANARIE and CFI. CANARIE's Research Software Program, which funds RS development and support, raises awareness of RS and RSE by promoting best practices, as well as leveraging digital research infrastructure resources. CANARIE's funding programs in RS[378] and RDM[379] have been in place for over 12 and 7 years respectively and are critical to many Canadian research platforms. This funding is designed both to encourage reuse of existing RS, and to promote awareness and monitor availability and (sustainability) of the tools. The CCF does not provide funding but provides decentralized and regional user support as well as platforms for conducting research on ARC clusters and research clouds.

CFI's cyberinfrastructure challenge included two competitions in 2015 and 2017, aimed at designing and developing research infrastructures involving multi-institutional consortia researchers, data scientists and SW developers. This funding included support for the CCF to enhance the capacity of Canadian researchers to conduct leading-edge computationally- and data-intensive research. These were CFI's first forays into supporting RS at a platform level (as opposed to hardware support), and in response to the growing awareness of the challenges of sustaining SW infrastructure. Another CFI award is the John R. Evans Leaders Fund (JELF)[380], which is a partnership[381] grant program with either SSHRC, NSERC, or CRC/CERC that serves the needs of individual researchers in terms of research infrastructure. There is also an option to have parallel applications jointly adjudicated at the same time or to follow up with a JELF.

Non-Canadian agencies have funded some RS efforts in Canada, including in the HSS. Several platforms have been successful in EU H2020 funding, including iReceptor[382] and CanDIG[383] through the joint EU Horizon 2020/CIHR call[384]. It is becoming fairly common that CIHR partnered with the Wellcome Trust and other international agencies to fund RS (e.g. FAIRsFAIR[385] project). The Mellon Foundation funds a substantial number of digital research infrastructure and software

---

378 https://www.canarie.ca/software/funding/lrss-call1/

379 https://www.canarie.ca/rdm/funding/call2/

380 https://www.innovation.ca/awards/john-r-evans-leaders-fund

381 https://www.innovation.ca/sites/default/files/Funds/JELF/nov2018/guidelines_for_completing_a_proposal_-_partnerships.pdf

382 https://gateway.ireceptor.org/

383 https://www.distributedgenomics.ca/

384 https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/sc1-bhc-05-2018

385 https://www.fairsfair.eu/news/rori-and-fairsfair-partner-fairware-initiative-funders-across-uk-europe-and-canada-join-forces

development projects worldwide[386,387]. Since 2012, with support from the Sloan Foundation[388], CLIR's (The Council on Library Resources) Postdoctoral Fellowship program[389] has provided support for both research data curation and SW curation by placing 49 scientists and social scientists at 38 host institutions across the United States and Canada. These fellows helped cultivate a more sophisticated understanding of data and software curation that contributes to a sustainable digital environment for research.

The Innovation Superclusters Initiative[390] is a recent strategy supported by the Government of Canada that would transform regional innovation and job creation efforts, by challenging Canadian private sector organisations of all sizes to collaborate with research institutions. In 2018, the Innovation Superclusters Initiative announced five Superclusters in Canada (Digital Technology[391], PIC[392], NGen[393], Scale AI[394], and Ocean[395]), with $950M over 5 years to promote large-scale collaboration among industry leaders, SMEs, and academics to develop and scale high-potential technologies in Canada. The Digital Technology supercluster has funding pipelines for specific RS components like repositories.[396]

Ontario's Early Researcher Awards (ERA) program is a discretionary, non-entitlement program that encourages new researchers who are working at publicly funded Ontario research institutions to build a research team, regardless of disciplines. Each award to a lead researcher provides up to $100,000 with an additional $50,000 to be matched from the researcher's institution and/or a partner organisation. These funds can be used to hire HQP, which can be applied to the area of software development.[397]

---

386 https://mellon.org/grants/grants-database/advanced-search/?amount-low=&amount-high=&year-start=&year-end=&city=&state=&country=Canada&q=infrastructure&per_page=25

387 https://mellon.org/grants/grants-database/advanced-search/?amount-low=&amount-high=&year-start=&year-end=&city=&state=&country=Canada&q=software&per_page=25

388 https://sloan.org/

389 https://www.clir.org/fellowships/postdoc/

390 https://www.ic.gc.ca/eic/site/093.nsf/eng/home

391 https://www.ic.gc.ca/eic/site/093.nsf/eng/00011.html

392 https://www.ic.gc.ca/eic/site/093.nsf/eng/00012.html

393 https://www.ic.gc.ca/eic/site/093.nsf/eng/00010.html

394 https://www.ic.gc.ca/eic/site/093.nsf/eng/00009.html

395 https://www.ic.gc.ca/eic/site/093.nsf/eng/00013.html

396 https://www.ic.gc.ca/eic/site/093.nsf/eng/00018.html: some project examples are Satellite-based Environmental Analytics, Fresh Water Data Commons, Precision Agriculture to Improve Crop Health, etc.

397 https://www.ontario.ca/page/early-researcher-awards-program-guidelines#section-0

8.2. Challenges & Opportunities

8.2.1. Lack of Targeted and Sustainable Funding

Canada distributes substantial funding for research and development (R&D), innovation and commercialization projects[398], but there is little in the way of focused, specific, and/or even long-term funding for RS. In comparison to other countries (Appendix I), Canada lags behind in support of an effective, cohesive and coordinated RS development community, with the exception of CANARIE's programs. By contrast, Australia and Europe are examples of jurisdictions that have invested large sums in a focused effort to build robust research infrastructures for all researchers. The fundamental reason is that the development of RS and platforms has not been a strategic focus of funding agencies outside CANARIE, and the previously-mentioned CFI initiative. Typically, CFI will fund national centralized compute infrastructure projects that propose the development of SW required for the infrastructure to work, but anything related to RS development after the award period is not eligible. This can leave researchers scrambling to find additional programs to fund maintenance and further development, which is especially problematic when a considerable community of practice relies on that SW infrastructure. This is partially because RS is typically developed as a byproduct of domain-specific research, and until recently has not typically been viewed as a "first class" output or a "DRI component". This represents a significant gap and challenge for Canadian researchers and diminishes Canada's leadership role in the development of Research infrastructure. The funding programs that do exist (Section 6.1.1.2) are either domain-specific (e.g. national and regional genome programs of Genome Canada and Genome British Columbia[399] (Genome BC)) or they fund continued and sustainable RS development indirectly, such as when a PI develops RS themselves, or uses research funding for this purpose. This is an unsustainable approach to building research infrastructure in Canada and there are significant opportunities to grow the RS community through stakeholders working together to coordinate funding in this area.

The CCF has created ARC infrastructure SW in many areas, such as job scheduling, software distributions system, cloud deployment automation, and the central database (CCDB)[400], but this is typically an organic process that balances necessity, available resources, and expertise within existing staff, rather than through a strategic long-term vision. More importantly, the equivalent of international investments that are earmarked for ARC infrastructure SW (e.g. those in the US or EU), have not been made in Canada. This gap could be potentially addressed by recognizing its significance, along with associated funding and support for Science Gateway-type platforms and underlying middleware that could trickle down to infrastructure providers like CCF who do not have a strategic mandate to support such projects.

Another related challenge that some funding agencies are struggling with is how to balance the support between generic and discipline-focused infrastructures. Most research infrastructures across all domains now include significant digital infrastructure in their core elements, and it can be difficult to differentiate the digital infrastructure provided by these and generic DRIs. As noted

---

[398] https://www.canada.ca/en/services/science/innovation/funding.html

[399] https://www.genomebc.ca/

[400] Compute Canada Database (CCDB): https://ccdb.computecanada.ca/security/login

by RISCAPE report[401], a point of differentiation is that research infrastructures in specific domains tend to mostly use the basic services of DRIs: “It is difficult to find research infrastructures that use higher level services developed in the e-infrastructure domain, but rather they make their own developments of services needed in the respective domains.” This was a challenge CANARIE attempted to address via focused funding and encouraging existing platforms to onboard research communities outside their disciplines.

While researchers generally understand the value of RS both as a fundamental part of the research process, and as an impactful output, funding agencies have been slower to recognize and support this academic work. For example, the Tri-councils generally do not allow RS purchase or development as an allowable expense[402]. This has led to a situation where the graduate students in the lab develop RS rather than projects being able to hire trained research software engineers. SSHRC previously had a dedicated funding stream for experimental tool development, including RS, but that has been discontinued. While it is possible to get funding to develop research software as a supporting element of a research project, there are very few funding sources designed to develop research software as the primary outcome of a research project. This gap in funding is an opportunity for other funding agencies to support the development of RS in a deliberate way that complements the Tri-council.

There are other administrative and operational roadblocks that researchers may face when working to build important RS tools. As mentioned in Section 9.1.2, despite the option of parallel applications with JELF awards, these count against institutional CFI envelopes, which means that the ability to apply is subject to internal prioritization processes. There is also a requirement for matching funds, which in some cases may mean that successful Partnership Grant[403] (PG) PIs cannot apply for JELFs for accompanying infrastructure, despite the theoretical ability to do so, and hence projects that have been peer-reviewed and successful in the most prestigious HSS research funding program in the country, are unable to apply for infrastructure to support that research. It would be beneficial to have an envelope-free stream of funding (as was the case with the CFI Cyberinfrastructure program) to promote inter-institutional collaboration in infrastructure development. Another challenge has been the CFI 40/40/20 matching model, which can be especially challenging with collaborative grants that span multiple provinces, and the range of approaches to provincial matching. In practice, RS prototyping and development is funded through a range of SSHRC funding streams including the Partnership Development and Insight[404] grants, but historically the absence of clear guidelines around eligibility of funding RS development through SSHRC has meant that outcomes were heavily dependent on the expertise on review panels.

---

401 https://riscape.eu/riscape-report/

402 https://www.unb.ca/research/ors/_resources/tri-agency_eligible_research_grant_expenses_and_supporting_evidence_requirements_6mar2017.pdf

403 https://www.sshrc-crsh.gc.ca/about-au_sujet/partnerships-partenariats/partnership_grants-bourses_partenariats-eng.aspx

404 https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/insight_grants-subventions_savoir-eng.aspx

RS sustainability in the SSHRC context is complicated by a different relationship between data production, dissemination, and reuse in which research, RS, RDM, and publishing overlap. Perhaps even more than in other fields, many digital artifacts, which is to say much data, in the humanities remain relevant beyond the initial research cycle in which they are created; they are subject to ongoing evaluation, interpretation, updating, enhancement, and reuse by other scholars. Scholarly editions of primary sources, digitized archives, and thematic research collections of scattered materials, for instance, constitute a form of publication that both advances the research of the scholars that produce it[405], and are often essential to the work of other researchers and students, and valuable to the general public, in and beyond the field for a generation or more. This means that making large, digitized collections or researcher-produced subsets thereof interoperable and computable, such as the Hathi Trust Research Centre's Data Capsules that allow computational processing of bodies of copyright-protected cultural content, is itself an important aspect of RS for a wide range of fields[406]. It also means that sustaining meaningful access to such humanities research datasets via the web, in addition to long-term preservation, is a key infrastructural component of supporting research in the humanities[407]. When HSS RS is not sustained, there is significant danger of data loss or loss of access[408,409], e.g. the endangered Métis Archival Project (MAP) site or the defunct ArtMob platform, and of the replication of significant labour to rescue data for active use, that is, online accessibility, if it is restored at all, as with the Fred Wah Digital Archive. Code itself is a cultural artifact that requires significant efforts to preserve and keep accessible, as attested by initiatives such as the UNESCO-supported Cultural Heritage Archive[410]. Complex interactive code artifacts such as e-literature and video games pose unique challenges that it is essential to address to make them susceptible to social and cultural analysis in the future.

Not only the availability of Cloud but effective RS management on Cloud and solutions for keeping content web-accessible are therefore an essential aspect of HSS DRI. The Endings project[411] promotes best practices to enable the "graceful degradation" (Nowviskie and Porter) of completed web projects, for instance through the use of static websites (e.g. using github.io). Such DH-specific methods must be promoted and offering managed versions of commonly used and

---

[405] Muñoz, Trevor, 2013. Data curation as publishing for the digital humanities. *Journal of Digital Humanities*, *2*(3), pp.14-22.

[406] Zeng, Jiaan, Guangchen Ruan, Alexander Crowell, Atul Prakash, and Beth Plale. "Cloud computing data capsules for non-consumptiveuse of texts." In Proceedings of the 5th ACM workshop on Scientific cloud computing, pp. 9-16. 2014.

[407] Terras, Melissa, James Baker, James Hetherington, David Beavan, Martin Zaltz Austwick, Anne Welsh, Helen O'Neill, Will Finley, Oliver Duke-Williams, and Adam Farquhar. "Enabling complex analysis of large-scale digital collections: humanities research, high-performance computing, and transforming access to British Library digital collections." Digital Scholarship in the Humanities 33, no. 2 (2018): 456-466.

[408] Edmond, Jennifer, and Francesca Morselli. "Sustainability of digital humanities projects as a publication and documentation challenge." Journal of Documentation (2020).

[409] https://robincamille.com/presentations/death_of_digital_scholarship/

[410] https://www.softwareheritage.org/

[411] https://endings.uvic.ca/

standards-based open-source RS such as Omeka or Mukurtu would make them more attractive and accessible to scholars and would offer even greater benefits in terms of metadata and data format consistency to support data sharing and reuse. Methods and tools in the SSHRC disciplines are so diverse that no single campus can support them all, so a pool of shared expertise nation-wide would enable researchers to make more informed choices from a broader pool of options. Many researchers currently adopt lower-barrier approaches such as WordPress sites, which are used to present both data that is of use to others and significant research results (e.g. Yellow90s[412]). Supporting both strategies would create links at early stages with researchers in ways that would advance best practices in research methods and RDM, ideally through local support teams tied into larger networks providing access to expertise in particular RS.

Sustainability is the capability to endure. For RS, sustainability means it will continue to be available and functional in the future, while the need for it exists or until it is superseded, which requires human effort during this full period. Because RS sustainability will lead to an overall benefit for the research community and better use of research funds, it needs to become a greater focus for funders. CANARIE recognized this need and chose to shift from funding new RS development to focusing on RS sustainability by growing and expanding previously funded projects and encouraging RS reuse. This of course led to less funding for new RS projects from CANARIE. This demonstrates the challenges that funders face in balancing funding new RS projects with ensuring that existing projects that are still useful are sustainable. Establishing such a funding balance is important to a healthy RS ecosystem.

Another challenge regarding funding of RS projects is around funding administration and management. Two of the key Canadian RS funding programs, CANARIE's RS and CFI's Cyberinfrastructure Challenge 1 program, fall outside of the normal Tri-council granting model. As a result, applying for and managing projects in these programs is outside the normal grant experience that researchers are familiar with. Both programs have significantly different application processes (e.g CFI grants are typically for large physical infrastructure rather than software), administrative, HR, legal implications (e.g. CANARIE's program is a claims-based contract model), and reporting requirements (e.g. both CANARIE and CFI have stringent reporting requirements) than projects funded by the Tri-council agencies. As a result, applying for and administering these projects carries a heavy administrative overhead that disincentivizes applications from even well-qualified applicants, and that even large research offices struggle to meet. It should be noted that these overheads are inherent in the requirements that the funders have with the federal government (e.g. CANARIE's Contribution agreement), and the funders have no choice but to impose these requirements. As such, opportunities exist to streamline the application and management processes around RS grants.

Another key disconnect in traditional research funding pipelines is that funders, and hence reviewers, look for direct and immediate novelty as one of the review criteria from the solicited proposals, which at some point in the lifecycle of important RS, is contrary to the needs of RS sustainability, where maintenance leads to use which leads to later novelty by those users, rather than immediate novelty, is the key. This novelty requirement provides a wasteful and counter-productive incentive for researchers to develop new RS rather than reuse existing components or add iterative enhancements. It also leads to competitive proposals from PIs that would be better

---

[412] http://www.1890s.ca/

served to work in a more collaborative environment where best practices for RS development are the norm, rather than the exception.

RS is often funded as infrastructure in a three- to five-year grant cycle, while the typical RS project should/would last much longer, especially where they serve a large community. Maintaining, adapting, and expanding RS are core requirements of RS sustainability. If funders plan to foster RS sustainability, they must provide funding for these purposes. It was noted that UK's EPSRC and the Biotechnology and Biological Sciences Research Council (BBSRC) have both run funding calls for the maintenance of software, namely Software For The Future II[413] and Bioinformatics and Biological Resources (BBR) Fund, respectively.[414]

Even good RS will not be used if no one is aware of its existence. Making RS discoverable and available will reduce duplication efforts, and foster RS reuse that allows both funding and researcher effort to be focused on research rather than RS development. If developers can articulate use cases for their RS, there is a greater chance that they will make a case for funding to maintain that RS. Equally important is designing the software with reuse outside of the originating group in mind, leveraging flexibility, plugin architectures, etc., which is perhaps the largest impediment to reuse by others. In commercial SW, more users usually mean more revenue, and thereby the ability to support those users, add new features, etc. However, in RS, without funding to hire dedicated HQP for user support, more users mean more time needed to support those users and less time to add features, undertake the actual research, etc. Such disincentives hinder the sustainable development of RS. A new approach to RS funding needs to recognize different types and phases of RS (e.g. experimental; emergent but production-level; established/enterprise), and devise appropriate evaluative mechanisms, metrics, and funding streams for each.

8.2.2. Technical

The RS ecosystem is varied and complex. RS includes a broad range of software, from systems to applications, modeling, Gateways, analysis, algorithms, middleware, and libraries; from highly developed packages with multiphysics algorithms and frameworks used by significant user bases to programming models and abstractions for data, hardware, and science, or very short and simple scripts, notebooks, and programs, written by researchers for their own use. When it comes to RS platforms, the challenge of domain-specific versus generic infrastructure development is complicated too. For instance, with some CANARIE funding programs, unless the proposed project was able to address the issue of how to take domain-focused RS tools and/or platforms and make it more generalizable it was not eligible for that funding program.

Over the last 15-20 years there have been enormous efforts to raise awareness of the importance of data to research, where interests in and investments into RDM and associated best practices have considerably improved how data are handled and preserved. However, SW is not data. Compared with data that are typically static, SW is an evolving entity that must adapt to the

---

[413] https://epsrc.ukri.org/funding/calls/softwareforthefuture/

[414] https://webarchive.nationalarchives.gov.uk/20200930155721/https://bbsrc.ukri.org/funding/filter/2019-bioinformatics-biological-resources-fund/

constant changes in its environment: if it does not evolve, SW will decay. Although research data is also susceptible to decay, the context is very different for RS.

SW relies on other software for its normal operation: these components are known as dependencies, and they are diverse, from the operating system, system libraries, to other necessary packages (e.g. a browser, Java Runtime Environment (JRE)). One way that software decays is via a change to any of the dependent software components that can affect the operation of the software, and the risk of this occurring is significant with large RS platforms. If RS is to continue to be of use, it must be sustained during its active lifecycle, rather than simply preserved.

Software development has been growing exponentially over the past half a century, and the tools and platforms that support it have been evolving at a fast pace. The number of original public source code elements doubles every 22 months, and original commits double every 30 months[415]. Software operates at a global level, and the accelerating pace of development in innovation, research and technology is incredibly rapid. For RS to remain relevant, it must evolve at a similar pace.

Another impact on the development of RS is the fact that research itself evolves at a rapid pace, putting even greater pressure on the actors in the RS ecosystem to respond accordingly. The response to the COVID-19 pandemic is a good example: many research programs in Canada and elsewhere pivoted quickly to respond to the myriad challenges presented by the crisis. For example, the iReceptor[416], QUEBEC COVID-19 BIOBANK[417] (BQC19), CanCoGen[418] and CanDig[419] platforms have responded quickly to make changes to their SW to accommodate research focused on COVID-19.

RS sharing is encouraged by some communities and promoted by funders, publishers and RS advocates, but typically not mandated. If RS is not widely disseminated or shared, it inhibits research transparency, reproducibility, and verification. Aside from best practice SW repositories like GitHub, there are no other RS repositories, catalogs, registries, or services in Canada except CANARIE's RS Portal[420] and the CCF framework. In addition, CCF offers a range of SW (ranging from 400 to 800 based on hardware platform types) across host sites for their users. The actual number of RS tools and platforms is difficult to quantify, but is many times what is readily available from national services. Scaling innovation and infrastructure to meet the range of requirements of diverse communities of practice is challenging without national coordination and support. Support

---

[415] Rousseau, G., Di Cosmo, R., & Stefano Zacchiroli. (2020). Software provenance tracking at the scale of public source code. Empirical Software Engineering, 1–30. https://doi.org/10.1007/s10664-020-09828-5

[416] http://www.ireceptor.org/, Scott et al. The adaptive immune receptor repertoire community as a model for FAIR stewardship of big immunology data, Current Opinion in Systems Biology Volume 24, December 2020, Pages 71-77, DOI: 10.1016/j.coisb.2020.10.001

[417] https://bqc19.ca/fr

[418] https://virusseq-dataportal.ca/

[419] CanDIG: Secure Federated Genomic Queries and Analyses Across Jurisdictions: https://www.biorxiv.org/content/biorxiv/early/2021/03/31/2021.03.30.434101.full.pdf

[420] https://science.canarie.ca/researchsoftware/home/main.html

for researchers to deposit their SW in accessible RS repositories that will catalogue, index, preserve, curate, and provide persistent identifiers, is critical to ensuring the sustainability of Canada's RS efforts. The Task Force on Best Practices for Software Registries of the FORCE11 Software Citation Implementation Working Group[421] presented a set of nine best practices that can help define the scope, practices and rules that govern individual RS registries and repositories.[422] The development of these best practices will only see benefit if it is accompanied by a sustainable investment strategy for the development of RS tools and platforms.

It is even more challenging to write RS for a complex ARC platform. Taking cluster-based programming for instance, both software and hardware go hand in hand when it comes to achieving ARC on a cluster. Programs must be coded to leverage the underlying hardware, while the traditional non-parallel architected programs must be re-written to perform well on a cluster. In addition, one needs to consider different approaches to parallel programming (distributed vs. shared memory approach), I/O handling, libraries that may not be supported in the future, etc. Despite these challenges, when RS for clusters is developed and managed in a sustainable manner, it can be maintained for decades, ultimately improving cost efficiencies as well as the advances in research. National services have a critical role here, whether offering interoperable and reusable RS across platforms, or supporting education and training.

A related challenge is how to integrate and coordinate the three components of ARC, RDM and RS in such a dynamic and evolving ecosystem. With an integrated and inclusive infrastructure (e.g., "one-stop shop"), researchers can benefit from producing, managing, and sharing their research outputs while enabling research reuse and reproducibility, as well as from facilitated collaborations and engagement across domains and institutions.

8.2.3. Skills, Knowledge, and Training

As outlined in the *Canadian DRI Environmental Scan* that was prepared for Summit 2014, "there is a significant unmet need for skills upgrading, training, and mentoring in the use of advanced computing, especially in disciplines that have not had extensive engagement in data-intensive research until recently."[423]

Some researchers know how to code, but few understand the wider set of skills that are needed to develop reliable, sustainable, reproducible, and reusable software, which generally fall under the scope of software engineering. On the other hand, not all researchers should have to become RSEs. Instead, it is necessary for a researcher to upskill appropriately for their level of involvement with software. Organisations like Software Carpentry address this problem by training researchers in research computing skills. Such training is also a key focus of CCF which offers a wide variety of training - programming, MPI, data analytics, ML, CUDA, scripting etc. - delivered in multiple formats (summer school, webinars, workshop etc.) to thousands of participants. This type of

---

[421] https://github.com/force11/force11-sciwg

[422] https://arxiv.org/abs/2012.13117v1

[423] The Leadership Council. (2014). *Canadian DI Environmental Scan: A Supplement to the Background Précis Document Provided to DI Summit 2012*, 7. Retrieved from http://digitalleadership.ca/wp-content/uploads/2014/01/Canadian-DI-Environmental-Scan.pdf.

training helps increase the general skill level of the research community, but if we are to significantly increase skills, ensuring that SW skills are provided at the very start of a research career is likely to ensure that these skills are used throughout that career. The lack of SW development skills in the education and research pipeline needs to be addressed and should be incorporated into doctoral training programs. Moreover, the lack of an organized RSE body makes it harder to provide skills training to that group. It is also important to achieve greater awareness of SW sustainability best practices, which can be facilitated by organisations (centralised or distributed) that act as focal points for expertise, not only to share knowledge and skills, but also to enhance networking and collaboration internationally.

8.2.4. Attraction and Retention of RS HQP

HQP in the context of RS refers to the professionals who are supporting and doing research, including researchers (including early career researchers, students and postdocs) and research support staff (e.g. sysadmins, research technical professionals (RTPs), software developers). The HQP generally considers 2 types: new HQP development during funded research; HQP who are members of an existing research team or infrastructure organisation, and already capable of conducting research activities and/or supporting research.

Among the RTP roles, the concept of an RSE is quite new, having only been more fully developed in the last decade. The birth of the term RSE can be attributed to the RSE movement in the UK, combined with a series of SSI-hosted workshops that highlighted the need for SW developers to have something more fundamental than simple role recognition. RSE is not traditionally considered a formal discipline, but as per the UK-RSE definition,[424] they are unique given a distinctive skill set that blends deep technical skills in SW development with an understanding of specific research environments and cultures.

Research efficiency and outcomes could be advanced if RSEs were embedded in research teams. Delegating responsibility for RS to an RSE also reduces the pressure on researchers to master yet another skill set, but it is challenging to attract and retain these RS HQP in academia. A common practice within academia is for PIs to encourage graduate students to undertake RS development, highlighting one of the more common pipelines for the development of RSEs. This approach is facilitated by a funding model that encourages the hiring of postdocs while discouraging the development of a more professionalized RSE resource. Secondly, the lack of a formal and sustainable career path for RSEs makes it difficult to recruit and retain such valuable staff. It can be argued that Canadian universities offer a positive and flexible working environment compared to that in industry, however the career path in academia can be less clear given that RSE positions are typically funding-dependent. This can lead to high turnover, and even when they stay, they are often in precarious positions. Take a two-year funded research project as an example: a RSE could be hired into this role quickly but find themselves in a dead-end position with no option for gaining recognition, reward or career advancement, because the typical university structure and academic culture rewards publications, not RS. In contrast, industry offers more financial rewards, team building, skills development, and promotion opportunities. This makes it difficult to hire and retain RSEs in academia, which in turn inhibits sustainability, as well as the proactive development of the RSE community. The challenges of recruiting RSEs in academia can be further exacerbated by the disparate needs of local HR and finance departments

[424] https://society-rse.org/about/

(e.g. local institutional employment contracts, salary grades and benefits, longevity and career path), and a university/funder bias to "traditional" researchers over RSEs, and the loss of HQP to industry. A robust recruitment and retention strategy in support of RS HQP, and especially RSEs, is needed.

CANARIE's RS programs signaled the start of the effort to develop an RSE culture within Canada, by hosting RS events (e.g. Canadian Research Software Conference), and by funding local research software support (LRSS) teams. In one example from the LRSS program after the two-year CANARIE seed funding, the McMaster University team[425] has secured funding directly from the institution, and McMaster is still engaging with other two pilot teams, the University of Regina, and Carleton University. All three teams have added members[426] (including RSEs, domain-specific research support staff, etc.) that CANARIE did not fund during the funding period, in response to high demand.

The development of the RS community is advancing slowly in Canada, and the RS community remains an informal one. An international model for retaining RS HQP is University College London, a pioneer in building RSE teams with support directly from the institution. This represents a new model for organising RS expertise, where research teams permanently employ RSEs. Their model is to fund a core capability and leadership team, reserving a portion of research funding overhead, with additional RSEs funded through paid services. This model also provides the opportunity for the RS group to develop and grow over time, although other models that have national support are emerging. For example, in the Netherlands, an approach involving sharing expertise, funding and responsibility across the DANS[427] and SURFsara[428] centres are being considered, acting as an "umbrella group that shares expertise across national organisations and campaigns for software sustainability at an EU level".[429,430]

The importance of high quality RSEs is becoming more prominent, within both institutions and funders (e.g. The EPSRC Software for the Future[431], NSF SI^2 funding[432]), and due to initiatives like those from CANARIE, the SSI, the Society of Research Software Engineering, Hidden REF[433], etc. As DRI moves towards exascale computing, we arrive at a pivotal point in the effective use of the next generation of technologies. Researchers have access to vast quantities of data, but they need help and the skills of the RSEs to enable them to use all that data. There

---

425 https://research.mcmaster.ca/research-resources/local-it/rhpcs/researchsoftware/

426 https://research.mcmaster.ca/research-resources/local-it/rhpcs/rhpcs-staff-directory/

427 https://dans.knaw.nl/en

428 https://www.surf.nl/en/research-ict

429 Hettrick, S. (2016). Research software sustainability: Report on a Knowledge Exchange workshop.

430 http://doi.org/10.1109/SE4Science.2019.00009

431 https://epsrc.ukri.org/research/ourportfolio/themes/researchinfrastructure/subthemes/einfrastructure/software/

432 https://www.nsf.gov/pubs/2013/nsf13511/nsf13511.htm

433 https://hidden-ref.org/

will be an increase in demand for RSE skills for institutions and research communities, both nationally and internationally.

Recognizing RSEs as a formal role is one thing, but funding and supporting career paths in it is quite another.[434] National implementations are defined by many different factors: how funding is structured; who are the key organisations; and the development of strategies for funding RSEs. With the significant demand for RSEs across domains of practice, and the potential for RS developed within one domain to have applications in other disciplines, there is a need to develop a comprehensive RSE funding model that is inclusive to RTPs and RSEs in all disciplines. Although Canada has a nascent RSE community in most regions of the country, there needs to be a greater focus on establishing both a national RSE community and a stable career path for RSEs within their host institutions and beyond.

This is a challenge for the DRI community as a whole and highlights the need to work together. It is important for funders to make it clear to all actors in the ecosystem that the contributions of RSEs and RTPs are encouraged and valued. Europe has recognized these new "technical meets traditional academic" pathways, and this RS development has substantial support. For example, France has a codified category of academic research support occupations (e.g. "research engineer"[435], associated with data analysis across all disciplines[436]). But these public service positions embedded in academic research institutes or projects have recently become more precarious positions within the framework of project-based research funding. Although historically, CNRS Portail des métiers IT [437] teams and laboratories have had career staff, these positions too have become less stable. In UK, RSEs and related fellowship programs are increasing[438]. An entity like the Turing Institute[439] has positions such as DevOps and Research software, while at the same time forges connections with other research partners. The signatories to the UK's Concordat to Support the Career Development of Researchers[440] set another good example of an approach intended to improve the employment and support for a rich profile of research careers, including RSEs. Following on these international examples, there is a significant opportunity in Canada to move from the current conceptual level of support and recognition for RSEs to a more systematic and strategic approach to aligning principles and policy to career incentives for RSEs.

---

[434] https://codeforthought.buzzsprout.com/1326658/7123159-funding-research-software-engineering

[435] https://www.enseignementsup-recherche.gouv.fr/cid23194/ingenieur-de-recherche.html

[436] https://metiersit.dsi.cnrs.fr/index.php?page=cartofamille&codeBAP=D&codeFamille=A

[437] https://metiersit.dsi.cnrs.fr/

[438] https://www.software.ac.uk/programmes-and-events/fellowship-programme

[439] https://www.turing.ac.uk/

[440] https://www.vitae.ac.uk/policy/concordat-to-support-the-career-development-of-researchers

8.2.5. Legal/Policy

Research software is international. Aligning policies (e.g. licensing management), standards (e.g. FAIR for RS[441]), and protocols (e.g. RS reuse) are critical for building excellence in RS, domestically and internationally. Unlike the best practices in RDM, where the efforts of Portage, RDC, and domain-specific communities have been working for decades, there is still much foundational work to be done in the RS context. Currently, Canada does not have well-developed policies, standards, and protocols to support researchers across disciplines in managing RS. This affects our ability to leverage the enormous potential of RS, but also influences our competitiveness as part of our international collaborations. There is a need to shape the national policy framework that guides the sustainable development of RS in Canada, while ensuring that Canada's framework is compatible with global practices.

Other jurisdictions, such as the EU, Australia and the US have invested in this area, and Canada can learn from the international RS community. There are some examples, such as participation in the International Research Software Engineering Leaders Workshop[442] and serving as a founding member of international Research Software Alliance[443]. In the last few years, several international RS community initiatives have emerged, such as RDA's Software Source Code IG (SSC IG)[444], Software Source Code Identification (SCID) WG[445], FAIR4RS WG[446], and the FORCE11 Software Citation Implementation Working Group (SCIWG)[447]. In addition to providing access to peer networks and opportunities to help shape international practice in RS, those international initiatives and communities also offer significant opportunities in terms of successful models: Canada could benefit from tools and resources that have already been developed.

8.2.6. Security and Privacy

Canadian research and innovation success are best supported by a high level of research integrity, and an open and collaborative research enterprise, across organisational types and jurisdictions, fostering research discovery and pushing the boundaries of research. At the same time, this open innovation foundation must be balanced by mechanisms that protect intellectual assets, discourage misappropriation of research artifacts and outcomes, and ensure responsible management of public funds. Behaviors that violate these foundational principles and values will jeopardize the integrity of the research and can even pose risks to the security of the research.[448] From this perspective, research security and research integrity are interrelated. Over the past decade in particular, individuals (domestic and international) and governments have exhibited

---

441 https://sorse.github.io/programme/workshops/event-016/

442 https://researchsoftware.org/

443 https://www.researchsoft.org/

444 https://www.rd-alliance.org/groups/software-source-code-ig

445 https://www.rd-alliance.org/groups/rdaforce11-software-source-code-identification-wg

446 https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg

447 https://www.force11.org/group/software-citation-implementation-working-group

448 https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97955.html

increasingly sophisticated efforts to exploit, threaten, and undermine research security and integrity, regardless of geographic location, size, and disciplines. Such behaviours include disclosure failures (e.g. funding; employment, affiliations and appointments; parallel laboratories; conflict of financial interests), inappropriate or exploitive behaviors (e.g. diversion of IP or other legal rights, breaches of contract and confidentiality), unethical and sometimes criminal activities (e.g. violation of ethical norms and federal agency policies; grant fraud; cyberattacks, data theft). Such behaviors can have substantial negative impacts on both individual researchers and research organisations, and threaten the integrity of the research enterprise, including distorting decisions about appropriate use of research funding; loss of funding; diversion of confidential or proprietary information and pre-publication data to foreign entities; reputational, career, and financial damage; and loss of public trust. In order to effectively address the challenges to research security and integrity, a coordinated, balanced, risk- and evidence-based approach must be established to protect Canadian interests and maintain the innovation ecosystem that has helped underpin Canada's global leadership in science and technology.

As concerns about cybersecurity continue to rise, it becomes an important and amplifying component of an organisation's overall risk management. And it is increasingly important to develop a robust national cybersecurity strategy beyond the scope of a single service provider or research university. Cybersecurity threats could apply to systems, data, people, assets, and capabilities, risking security, economy, reputation, public safety, and health. Some identified risks include protecting intellectual property, sensitive data and personal information; developing and implementing standards and policies for cybersecurity; awareness training; and privacy. Countering security threats effectively requires an ongoing effort to understand and manage cybersecurity risks, identify and implement measures to improve data security, prevent internal breach prevention, establish incident response processes, and maintain compliance with relevant requirements. That is why cybersecurity is becoming a major consideration for many RS developers, whether it is protecting critical services via appropriate safeguards (e.g. writing high-quality code that adheres to standards and best practices; identity and access management; strengthening cryptographic standards and validation; promoting trustworthy platforms), securing emerging technologies (e.g. IoT cybersecurity[449]), detecting and identifying the occurrence of cybersecurity events, responding quickly to cybersecurity incidents, restoring and recovering impaired capabilities or services, or maintaining plans for resilience.

Even more than other areas of SW development, the rapid pace of new types of cybersecurity threats means that most RS teams are either not well-positioned to understand or to respond to such threats. This is particularly important with critical infrastructure systems like Science Gateways, which may be particularly susceptible to cybersecurity threats because they are designed to interact with external users. Although this may be less of an issue for a researcher writing RS on an ARC cluster, as the environment and the security aspects in many cases are under the control of the system operator, not the developer or user of most typical application software, cybersecurity still needs to be taken into consideration. Skilled RSEs play a key role here, as they are more likely to have training in using security best practices in their software development.

---

[449] https://www.nist.gov/blogs/cybersecurity-insights/rounding-your-iot-security-requirements-draft-nist-guidance-federal

In response to the need identified by Canada's research and education community for national coordination and alignment of cybersecurity efforts, in 2020 CANARIE launched a new national program, the Cybersecurity Initiatives Program (CIP)[450], to fund cybersecurity initiatives that will strengthen the sector with advanced technologies, improved processes, and broadened expertise across Canada's Research and Education Sector. In addition, the recent release of the "Recommended Practices for Strengthening the Security and Integrity of America's Science and Technology Research Enterprise"[451] from the US National Science and Technology Council (NSTC)[452] offered recommendations for research organisations and is complementary to National Security Presidential Memorandum 33 (NSPM-33[453]), which directed federal departments and agencies to act to protect federally funded research. The NIST Cybersecurity Framework[454] also serves as a useful resource to help research organisations establish and maintain effective cybersecurity measures. NIST is developing a framework for RDM based on the Cybersecurity Framework.

The health data context - the need to protect patients' privacy versus the utility of that data - has special and complex issues of security. Data not shared is a missed opportunity to advance research in the interests of all, but inappropriate disclosure of data can compromise the interests of the individual. Research platforms such as Science Gateways that deal with health data require specific skill sets and knowledge that span best practices of developing secure software, understanding data privacy policy, and being able to map that to a RS solution. For example, The NIST Privacy Framework[455], provides a set of tools[456] and procedures for assessing security and privacy controls that are designed to protect sensitive but unclassified information by government, industry, and academia in support of various federal programs. RSEs need to be able to incorporate such policy frameworks into their RS solutions.

Institutions are at a turning point for how they respond to security risks, as the government pushes robust security responses, especially in human-focused research. One traditional approach, often embedded in institutional policy frameworks, is to control what researchers can do with respect to SW, through purchasing policy or the hiring of SW developers. However, this type of approach will not always have the expected outcome with respect to response to the associated policy. This approach compromises institutional and national security, making it incumbent upon the various actors to find ways to meet researchers' needs in a proactive and effective manner.

Security is too often erected as a wall that prevents users from accessing research outputs, leading to a tension in the research enterprise that must be carefully managed. With the

---

[450] https://www.canarie.ca/cybersecurity/cip/

[451] https://trumpwhitehouse.archives.gov/wp-content/uploads/2021/01/NSTC-Research-Security-Best-Practices-Jan2021.pdf

[452] http://www.whitehouse.gov/ostp/nstc

[453] https://trumpwhitehouse.archives.gov/presidential-actions/presidential-memorandum-united-states-government-supported-research-development-national-security-policy/

[454] https://www.nist.gov/cyberframework

[455] https://www.nist.gov/privacy-framework

[456] https://csrc.nist.gov/publications/detail/sp/800-172/final

appropriate implementation of best practice, awareness, training and education, and risk management, research outputs can be protected in an appropriate way, while at the same time preserving the open and collaborative approach to research critical to Canada's world-leading research and development.

8.2.7. International Collaboration

International RS collaborations are currently quite challenging. With limited exceptions, collaborative funding from national funding agencies can only be distributed to institutions within that country. For example, non-Canadian institutions are not typically able to directly receive funding from Tri-Council programs, or the percentage of available funds for international HQP need to be justified and may be capped at a very low level (e.g. CANARIE programs). Although it is possible to fund international researchers from Tri-Council funding, formal collaborations through these funding programs are challenging to establish. The same holds true for other international programs. For example, the Horizon Europe funding program does not in general allow Canadian institutions to directly receive funding, and funding researchers outside EU eligible countries is typically not allowed except under rare circumstances. As such, international RS collaborations are often loosely coupled, typically involving national funding of national components with the international collaboration done through informal agreements between research collaborators.

There are several notable exceptions to this. In 2018 the EU Horizon 2020 and CIHR agencies launched an international funding round focused on health data projects[457,458]. This funding program has supported six large international RS projects with substantial contributions from institutions across Canada.[459] In addition, some national funding programs allow for international collaborators to receive direct funding (e.g. the US NIH allows Canadian universities to receive funding) and there are some international funding programs for RS (e.g. Chan-Zuckerberg Open Source for Science[460]).

Finally, some domains are heavily international in their RS focus, with a history of international collaboration on RS. For example, the particle physics community has a history of international collaboration around RS development, with Canadian particle physicists both heavy users as well as significant contributors to this RS development.

Although establishing international collaborations on developing RS (through research funding) is currently challenging in Canada, there are several opportunities to work with stakeholders to address the issues and facilitate the use of and expand on some of the successes listed above.

8.2.8. Incentives and Metrics

Incentives related to metrics, funding, reward/recognition, career progression, and evolving scholarly norms are critical in catalyzing RS engagement. Generally speaking, researchers in

---

[457] https://cihr-irsc.gc.ca/e/50691.html

[458] https://cordis.europa.eu/programme/id/H2020_SC1-BHC-05-2018

[459] H2020 SC1-BHC-05-2018 projects

[460] https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/

disciplines and areas of inquiry from STEM to areas of the SSH that rely on compute-intensive, data-driven research develop and reuse RS on a regular basis, and view RS as valid research outputs. Researchers in other disciplines may not have the same view of RS as a primary output. There is an opportunity for all stakeholders to enact policies and programs that recognize the important role RS plays in achieving research outcomes, in all disciplines. In addition, programs and incentives are needed to encourage researchers and research support professionals (e.g. RSEs, RTPs and data stewards) to adhere to software development best practices, for increasing core SW expertise, and for training of research support professionals.

Capturing metrics about quality, level of development, and effort allocated to a specific RS project is another challenge, including the differences between communities of practice. In the humanities for example, there is a hesitation to cite digital versions of artifacts because it's considered more authentic to cite physical objects. These and other considerations suggest that there is an opportunity for policy makers to support initiatives that develop research assessment systems (e.g. the DORA principles that is being used by research agencies such as the Wellcome Trust[461]) that reward software outputs alongside publications, data and other research outputs, as well as ensuring there are proactive responses when these are not implemented. Traditional publication metrics such as citation-based metrics, should be considered complementary to emerging metrics (e.g. Altmetrics[462], GitHub forks, commits, media coverage) as a way of measuring RS impact appropriate to all actors in the research ecosystem. Advocating and creating appropriate metrics for RS activity would facilitate the culture change that is needed to realize the promise of open science.

8.2.9. Research Culture

A substantial intellectual and financial effort has been invested in the development of infrastructures, tools, and best practices that enable Open Science. Despite these investments, it is still not part of most researchers' routines. The impact of Open Science practices on software and reproducibility requires substantial knowledge and skills from the SW development team. For example, making RS available requires a special skill set, and one different from the dataset context. This presents an opportunity for collaboration between funding agencies and other stakeholders to create new incentives and metrics that strike a balance between Open Science and the traditional publish-or-perish culture.

In the research context, IP (data and outputs such as scholarly publications or creative works) is the currency of academic achievement alongside support for its production such as grants; however, IP takes very different forms in different disciplines. Data have traditionally not been widely shared, as they are not protected by copyright, and controlling their distribution has been how researchers leverage the value of those assets. There are also cases, for example The Canadian Peoples census data[463], in which a dataset cannot be fully open but can be shared if adequate personnel support is available. Conversely, the inadvertent relinquishment by academic authors of copyright to publishers impedes the sharing and reuse of research publications as open data. It is thus important to develop a culture in which sharing information is explicitly discussed

---

461 https://wellcome.org/grant-funding/guidance/data-software-materials-management-and-sharing-policy

462 https://www.altmetric.com/

463 https://thecanadianpeoples.com/

and encouraged, and this process is best fostered by individual research communities, since norms vary by discipline. RS has a role to play in supporting such culture change.

To truly exploit the new DRI, both technological and cultural factors need to be given equal weight to advance the research that is critical to building and sustaining Canada's economic and social prosperity. To achieve that goal, a set of best practices leading to barrier-free, open access to RS need to be developed, and maintained. A transformation to open science-driven RS culture makes scientific processes more efficient, transparent, and responsive to societal challenges, and makes scientific knowledge more easily accessible. Researchers often overlook the fundamental contribution that software makes to the reliability and reproducibility of their scientific results; publishers overlook the need to identify software as a vital part of the publication process; funders overlook the need to make funding available for maintaining software in a sustainable manner and overlook the growing need to secure software experts on research projects. Research institutions and organisations overlook the need to build SW expertise in research teams; policymakers overlook the importance of software to research by skipping the application of appropriate impact metrics. A sustained campaign of awareness-raising is required across the research community and for all actors, to convince research stakeholders that RS is a valuable research output in line with the investment it receives and the research it facilitates.

A final cultural observation is that active researchers may not know of others, in their discipline or elsewhere, who are undertaking similar or complementary efforts in the RS context. There are many reasons for this, but lack of recognition of RS as scholarly work, the gap in understanding between RS development and research needs, or the lack of funding for RSEs and related staff, and the lack of national RSE community initiatives are examples that can have a significant impact. Initiatives like the International Research Software Engineers[464] and CANARIE's RS conference and their LRSS program present opportunities to grow and expand the RSE community in Canada.

8.2.10. Government Policy/Strategy

There is an opportunity for federal and provincial research organisations (e.g. NRC and NAPHRO) to develop a strategy for working more collaboratively with their university partners, and across governments. The large number of users & communities, the complexity of clarity on their roles and responsibilities (international, national, regional, and local), the number of jurisdictions that are involved, and the diversity of requirements with which they must align their efforts, highlight the challenges in this task. Although national coordinated leadership for RS is emerging, efforts to crystalize RS communities have not been funded adequately or mandated formally. These two factors have made it challenging to undertake strategic and coordinated planning. When combined with the lack of coordination of investment in the context of RS made it difficult to develop the shared policies, processes, protocols, best practices, and standards that are so essential.Models that can inform such coordination exist; an example of governance and engineering bringing RS and data from multiple countries into a coherent framework is the EU Open Science Cloud.[465]

---

[464] https://researchsoftware.org/

[465] https://eosc-portal.eu/

8.2.11. EDI and Representation

As discussed previously, developing an effective approach to EDI can be problematic in the research software context. Despite a strong base of Computer Science programs in Canada, there are no specific academic programs in RS engineering in the country, leading to a lack of SW development skills in the research pipeline, but also to a lack of promotional contexts for RSEs. The lack of EDI in the research software context stems from various factors, including significantly fewer women and BIPOC representation entering Comp Sci programs and/or SW-related careers, as well as the generally narrowing path to tenure for women researchers[466]. This indicates that we need to build a diverse workforce by raising awareness among various stakeholders (individuals, research organisations, and policy makers) about challenges and opportunities related to EDI for the benefit of the RS profession and creating/advancing programs that aim to expand opportunities and eliminate barriers for underrepresented communities interested in RS[467]. There is some work in this area, but not nearly enough: The US-RSE Diversity, Equity, and Inclusion Working Group (DEI-WG) has been formed and is actively working on organizing DEI initiatives and related topics, engaging with DEI Technology groups such as Women in HPC[468] and Code 2040[469], as well as developing guidance on how to write inclusive RSE job descriptions to ensure they attract a diverse audience.

A more recent but related area is whether open science and reproducible science are sufficiently inclusive for some in the community that may be harmed by the control or protection of toolchains[470]. This aspect of security addresses the issue of control over data and code as a power differentiator. This involves challenging aspects of research culture and risk, which can be partly addressed by creating systems and workflows in research communities or teams that protect vulnerable populations and addresses systemic patriarchy, colonialism and racism ultimately enabling an equity, diversity and inclusivity (EDI) dimension in the RS context. Guidance on best practices towards reproducibility and open science is just one way to facilitate advances and EDI and benefits for trainees and junior researchers.

9 Next steps

Readiness to respond to current challenges and potential opportunities requires a thorough understanding of the current state and landscape, in agreement with short- and long- term goals and objectives. As this report summarizes the current state of the RS landscape in Canada to support developing a common understanding among the Alliance's members of the breadth and complexity of stakeholder engagement in this field, it also serves as a basis from which the Alliance can set a path forward for national strategy for RS in Canada. Findings and observations

---

466 https://nplusonemag.com/issue-34/essays/sexism-in-the-academy/#fn1-11214

467 https://developerrelations.com/dev-rel/mentored-sprints-for-diverse-beginners

468 https://womeninhpc.org/

469 http://www.code2040.org/

470 https://en.wikipedia.org/wiki/Toolchain

in this document, alongside the RDM and ARC Current State Assessment publications, are meant to help support the researcher needs assessment process, the Alliance's new service delivery model and funding model development, and DRI strategic plan processes.

9.1. Researcher Needs Assessment

The Alliance has consulted the Canadian research community to assess current DRI tools, services, and support; identify and address the ideal future state of DRI in Canada; and how the Alliance could achieve such a state from ARC, RS and RDM perspectives, respectively. The comprehensive consultation process consists of four steps: Step 1: Call For White Papers[471] on Canada's Future DRI Ecosystem (completed); Step 2: Call For Current Documentation (completed); Step 3: Online Survey (completed); Step 4: Virtual Town Hall Discussions (completed). The Researcher Council[472] oversees the researcher needs assessment process that will engage a wide range of researcher communities and disciplinary associations and provide independent advice to the Alliance on matters related to the delivery of services and programs for the research community.

Following the needs assessment process, outcomes will be integrated with findings from the

ARC, RS, and RDM current state reports that make recommendations to support the Alliance's strategic planning effort.

9.2. National Service Delivery Model and Funding Model Development

In collaboration and consultation with Deloitte and DRI partners, the Alliance will refine a new service delivery model that defines national, regional, and local services for RDM, ARC, and RS, including expected service levels, new funding models, transition roadmaps, and roles and responsibilities of stakeholders.

9.3. DRI Strategic Plan for 2022-24

The Alliance will present a national strategy and vision for ARC, RS, and RDM, integrating findings from its assessment and outreach activities. The strategic plan will include a roadmap for transforming the ecosystem from the pillarized current state, where ARC, RS, and RDM are treated as separate entities, into a more integrated desired future state where research is supported by robust DRI across its lifecycle.

---

[471] https://engagedri.ca/canadian-digital-research-infrastructure-needs-assessment/white-paper-submissions

[472] https://engagedri.ca/researcher-council

Acronyms and Glossary

This acronym and glossary table is intended to aid readers in understanding the meaning of selected terms as they are used in this document.

Full name	Acronym	Full name	Acronym
Leadership Council for Digital Research Infrastructure	LCDRI	Research Software Engineers	RSEs
digital research infrastructure	DRI	Canadian Research Software Conference	CRSC
data management	DM	Porous Materials Engineering and Analysis Lab	PMEAL
advanced research computing	ARC	McGill Centre for Integrative Neuroscience	MCIN
Research Software	RS	Engineering and Physical Sciences Research Council	EPSRC
New Digital Research Infrastructure Organization	NDRIO	Common Lab Research Infrastructure for the Arts and Humanities	CLARIAH
Working Group on Research Software	RSWG	Vice-Presidents Research	VPR
Pacific Institute for the Mathematical Sciences	PIMS	Information technology departments	ITD
Software	SW	Chief Information Officers	CIO
Earth Science Division	ESD	National Alliance of Research Health Funding Organizations	NAPHRO
Earth Science Data Systems	ESDS	Compute Canada Federation	CCF
Interactive development environments	IDEs	Canada Foundation for Innovation	CFI
Department of Energy	DOE	Canadian Institutes of Health Research	CIHR
Office of Scientific and Technical Information	OSTI	Natural Sciences and Engineering Research Council	NSERC
free and open-source software	FOSS	Social Sciences and Humanities Research Council	SSHRC
intellectual property	IP	open-source Program Offices	OSPOs
National Initiatives for Open Science in Europe	NI4OS Europe	Ontario Centres of Excellence	OCE
Contribution License Agreement	CLA	International Research Software Alliance	ReSA
HPC Container Maker	HPCCM	Research Data Alliance	RDA
Astrophysics Source Code Library	ASCL	Journal of Open Research Software	JORS
International Conference on	ICSE	Journal of open-source	JOSS

Software Engineering		Software	
Highly Qualified Personnel	HQP	open access	OA
National Research and Education Network	NREN	Canada Research Continuity Emergency Fund	CRCEF
Software Sustainability Institute	SSI	Canada Emergency Wage Subsidy	CEWS
Research Data Management	RDM	Canadian International Innovation Program	CIIP
National Data Services Framework	NDSF	National Research Council of Canada Industrial Research Assistance Program	NRC IRAP
Canadian University Council of Chief Information Officers	CUCCIO	Partnership Development Activities	PDAs
Canadian Higher Education Information Technology	CANHEIT	Requests for Proposals	RFPs
Research & High Performance Computing	RHPC	small and medium-sized enterprises	SMEs
Canadian Advanced Network for Astronomical Research	CANFAR	Canadian Statistical Sciences Institute	CANSSI
National Research Council Canada	NRC	Smart Prosperity Institute	SPI
Indigenous Knowledge Social Network	SIKU	Michael Smith Foundation for Health Research	MSFHR
New Frontiers in Research Fund	NFRF	Canadian Partnership Against Cancer	CPAC
Canada Research Coordinating Committee	CRCC	Dystonia Medical Research Foundation Canada	DMRFC
Tri-agency Institutional Programs Secretariat	TIPS	Graham Boeckh Foundation	GBF
Canadian Space Agency	CSA	Integrated Youth Services	IYS
Flights and Fieldwork for the Advancement of Science and Technology	FAST	nternational Alliance of Mental Health Research Funders	IAMHRF
Space Technology Development Program	STDP	Global Affairs Canada	GAC
announcement of opportunity	AO	Grand Challenges Canada	GCC
research and development	R&D	Bill and Melinda Gates Foundation	BMGF
Innovation, Science and Economic Development Canada	ISED	Declaration on Research Assessment	DORA
Innovative Solutions Canada	ISC	Ministry of Research, Innovation and Science	MRIS
European Space Agency	ESA	Ontario Research Fund – Research Excellence	ORF-RE
Canada First Research Excellence Fund	CFREF	Ontario Institute for Cancer Research	OICR
Canada Research Chairs Program	CRCP	Scholarly Infrastructures for Research Software	SIRS
Canada Excellence Research	CREC	FAIR For Research Software	FAIR4RS WG

Chairs	
Networks of Centres of Excellence	NCE
transparency, reproducibility, ethics, and effectiveness	TREE
infrastructure as a service	IaaS
platform as a service	PaaS
software as a service	SaaS
Infrastructure as Code	IaC
Internet of Things	IoT
Quality of Service	QoS
Edge Computing	EC
Virtual Machine	VM
Software-Defined Networking	SDN
Network Function Virtualization	NFV
Quantum computing	QC
information and communication technology	ICT
Master of Engineering Leadership	MEL
Centre de recherches mathématiques	CRM
Atlantic Association for Research in the Mathematical Sciences	AARMS
John R. Evans Leaders Fund	JELF
humanities and social sciences	HSS
The Council on Library Resources	CLIR
Protein Industries	PIC
CCF central database	CCDB

Working Group	
General Data Protection Regulation	GDPR
Partnership Grant	PG
Biotechnology and Biological Sciences Research Council	BBSRC
Bioinformatics and Biological Resources	BBR
Java Runtime Environment	JRE
research technical professionals	RTPs
local research software support	LRSS
RDA’s Software Source Code IG	RDA SSC IG
Software Source Code Identification	SCID
FORCE11 Software Citation Implementation Working Group	FORCE11 SCIWG
Cybersecurity Initiatives Program	CIP
US National Science and Technology Council	NSTC
National Security Presidential Memorandum 33	NSPM-33
Equity, diversity and inclusivity	EDI
Diversity, Equity, and Inclusion Working Group	DEI-WG

Appendix A: Science Gateways Examples in Canada

Science Gateway Name	National vs. International	General vs. Domain-specific	Branch of Science	Service Catalogue
2i2c	International	General	Interdisciplinary	Computational tools; Processing & analysis; Training & support
ATLAS-Canada	International	Domain-specific	Natural sciences/physical sciences/particle physics	Compute ; Data management; Processing & analysis; Storage
CANFAR	National	Domain-specific	Natural sciences/physical sciences/astronomy	Data management, Compute, Processing & analysis, Sharing & discovery, Storage, Training & support
CBRAIN	National	Domain-specific	Support Activities/Infrastructure development; Medical and Health Sciences/biomedical informatics; interdisciplinary	Data management, Compute, Security & operations, Processing & analysis, Storage, Training & support
CWRC	International	Domain-specific	Social sciences and humanities	Processing & analysis; Training & support
Érudit	International	Domain-specific	Support Activities/Digital archives	Data management, Security & operations, Sharing & discovery, Storage
GENAP	National	Domain-specific	Medical and Health Sciences/biomedical informatics; Genomics;	Data management; Networking; Processing & analysis; Security & operations; Sharing & discovery; Storage; Training & support
iReceptor	International	Domain specific	Health sciences, Immunology	Compute; Data management; Processing & analysis; Sharing & discovery; Storage; Training & support
MERIDIAN	International	Domain-specific	Natural sciences/underwater acoustics	Data management, Compute, Security & operations, Processing & analysis, Sharing & discovery, Storage, Training & support
Oceans 2.0	International	Domain specific, oceanography	Natural sciences/Earth & related environmental; Ocean management, disaster mitigation, and environmental protection; Physical, chemical, biological, and geological oceanography	Data management; Processing & analysis; Security & operations; Sharing & discovery; Storage
PanGeo	International	Domain-specific	Earth sciences/ geoscience	Computational tools; Processing & analysis; Storage; Training & support
PAWS	International	General	Interdisciplinary	Computational tools; Processing & analysis
SNOLab	National and International	Domain-specific	Natural sciences/physical sciences/ Astroparticle Physics	Processing & analysis

Syzygy	National	General	Interdisciplinary	Compute, processing & analysis Sharing & discovery, training & support,
Voyant Tools	International	General	Interdisciplinary, text analysis	Sharing & discovery; Processing & analysis; Training & support

Applendix A lists several examples of Science Gateways. More examples could be found via CANARIE's Research Software Platform Registry[473].

Appendix B: RS Support and Services Provided into Daily Research Activities

473 https://science.canarie.ca/researchsoftware/platforms/list/main.html

	Research Lifecycle							Data	Compute	Software	
	Plan	**Create**	**Process**	**Analyze**	**Disseminate**	**Preserve**	**Reuse**	**RDM Tools**	**System performance and monitoring tools**	**non-ARC/non-RDM Tools**	**Science Gateways Tools**
RS Activities, Support and Services	Idea discovery & generation	Initial design, creation & development of RS			Grant permission by the contributor(s) to release the RS		Re-discover/re-search RS	Data Management Plan (DMP) tools	Languages	Transfer	Integration
	(Sustainable) RS management planning & conceptualizing	Deploy/use/modify/customize RS			(Re-)consider choosing a RS licensing		Acquire, collect, receive and assemble resources: RS & dependencies, related data, publication and documentation, (computational) research workflow, etc.	Generate/produce/create/simulate data and related metadata		OS	Middleware
	Identify, acquire, collect, receive and assemble resources (including funding and people) into teams and communities	Describe/document RS (e.g., README file)			Appraise & select RS to be released		Set up running RS environment (e.g., containers, notebooks, cloud-based reproducibility environment)	Collect data	Libraries	Analysis	Toolchains
		Version control			Prepare metadata, documentation & instructions for RS (usage)		Reproduce, verify and validate results to maintain scientific integrity	Clean/process/analyze (modeling, simulation, data analysis/visualization)/ manipulate/transform data	Scripts	Desktop VMs	Web technologies
		RS testing, debugging & integration			Organize & store/backup RS		Integrate RS into new research	Data wrangling & visualization	Pipelines	Visualization	APIs
		Organize & store/backup RS			Version control			quality assurance/control	Large-scale numerical modeling/simulation packages/frameworks (model + data)	Pre-processing	

	Research Lifecycle							Data	Compute	Software	
	Plan	**Create**	**Process**	**Analyze**	**Disseminate**	**Preserve**	**Reuse**	**RDM Tools**	**System performance and monitoring tools**	**non-ARC/non-RDM Tools**	**Science Gateways Tools**
		Recognize contributions to and of software			Choose a publishing platform for RS / register RS in a registry	Choose a reliable long-term platform for RS / register RS in a registry		Describe, document, interpret data	Frameworks	Post-processing	
RS Activities, Support and Services		Respect existing SW licensing / choose a RS licensing for new developed RS			Ingest/deposit/share/release RS	Ingest/deposit RS for preservation		Data integration	Data transfer		
		Maintainance & ongoing testing/debugging of RS						Organize data	Arrange storage		
								Store & back up data (tools and platforms)	Hardware		
								Version control	Advanced/streaming data management, processing, computing & analysis		
								Share/publish/preserve data into repositories/registries/platforms	Pre-/Post-processing		
								FAIR data	Visualization		
								Reuse data	OS		
								Research reproducibility	File Systems		
								Access data	(Commodity) clouds		
								(Re-)discover/search data (tools and platforms)	Scheduler		

	Research Lifecycle							Data	Compute	Software	
	Plan	**Create**	**Process**	**Analyze**	**Disseminate**	**Preserve**	**Reuse**	**RDM Tools**	**System performance and monitoring tools**	**non-ARC/non-RDM Tools**	**Science Gateways Tools**
RS Activities, Support and Services								Data curation/preservation /archiving (data selection, appraisal, conversion, etc.)	Virtualization Desktop infrastructure (VDI)		
								Sensitive data (de-identification)	Integrate middleware		
								Secure & protect data (tools and platforms)	System performance and monitoring tools		
								File systems, OS	Benchmarking and scalability tools		
								Transfer data	Debugging tools		
								Measure scholarly impacts (views, downloads, citations, Altmetrics, etc.)	Cluster systems management tools		
								Migrate/transform data from the original into different formats and through suitable media	Software repositories		
								Digitization			
								Science Gateways			
								Automation: reusable workflows (e.g., YesWorkflow, Taverna, Kepler)			

Appendix C: Archetypes of Users & Communities in the RS Ecosystem

By job roles/archetypes	Description	Related types, roles, and/or examples
Researchers	Someone who conducts research	Faculty, students, postdocs, non-faculty researchers (e.g., Lecturers, Research/Lab Scientists, Research Associates/Fellows/Assistants/Staff) in academic sector; researchers in the Research & Development (R&D) department of the non-academic sector
Librarians	Someone who works in the library, providing access to information, instructions and information literacy	Metadata Experts, curators/archivists, subject librarians, cataloger, support/QA, copyright librarian
Sysadmins	Someone who is responsible for the planning, execution, overseeing, monitoring, support, operation, maintenance and configuration of the information technology systems, single/multiple computers and/or servers	Gateways operators, SW/application support, facility staff, system engineer, IT security support, licensing/copyright staff
Developers	Someone who writes code, software and/or computer programs	Professional developers (e.g., Research Software Engineers (RSEs)), analysts, self-taught/trained developers (e.g., researchers, postdocs, grad students), open-source communities (Numpy, Pandas, Jupyter, ...); software developers (software engineers, analysts, etc.) in researchers in the in non-academic sector
Research Group	A group of researchers who work together on different aspects (planning, executing, overseeing, educating, etc.) of a project	Project managers, user training/support, grant managers
Commercial Vendor Partners	A commercial entity or enterprise that is part of the alliance supplying services or goods	Commercial developers, R&D developers; data providers; interns, affiliated scholars, NGO participants,
Community Support	The support that is offered and/or obtained from a dedicated community	Mailing lists, forums, open-source communities (Numpy, Pandas, Jupyter, ...)
Academy + Nonacademy Research Collaboration Partnerships	The alliance between academic and nonacademic entities	Interns, affiliated scholars, NGO participants, citizen scientists/researchers
Funders	Funding agencies, coropirations and foundations	public funders, private funders (e.g., Wellcome Trust)
Journals & Publishers	A person, periodoical publication, compnay or organization that prepares, distributes and issues academic research and scholarship	Elsevier, Journal of Open Research Software (JORS), Journal of open-source Software (JOSS), Software Impacts, SoftwareX
Other stakeholders	Other archetypes than listed above	Government, academia, industry & commerce

Appendix D: Types of RS

Alternative ways of categorizing RS are based on its characteristics and features. For instance, RS may be categorized

- by functions: e.g. application software, system software, programming tools, software service, research platform
- by layers[474]: e.g. platform & infrastructure, tools & libraries
- by licensing status: open-source, closed-source, hybrid
- by publication status: published, unpublished
- by instantiation status: proof of concept, actual functional operational product
- by cost: e.g. free, commercial
- by the software distribution mechanism: e.g. source code, binary executable, package, container, virtual machine image, service
- by disciplines: general, domain-specific
- by research methods: quantitative, qualitative
- or by purposes of processes in the research lifecycle: planning, analysis, computation, visualisation, transfer, storage, publishing, curation and preservation, discovery

In practice, a piece of RS may not always fit neatly into a single type. In that case, the RS has hybrid types. For example, the RS may be composed of both open-source and closed-source components, or even of both published and unpublished elements. Another example is when there is source code and something else, e.g. a service or an executable that is built from that source code, but not by the user. In either case, it is important to be able to identify each file or component uniquely and independently to make sure which metadata and SW type apply to each component: license, version, creators, etc. There could also be a hybrid of RS and other research objects, such as a Jupyter notebook or a computational workflow or script that is published within mixed-type objects known as Research Objects[475] or Research Compendia[476] that represent computational pipelines with associated provenance information.

[474] https://hal.archives-ouvertes.fr/hal-02117588/document

[475] https://zenodo.org/communities/ro/search?page=1&size=20

[476] https://biostats.bepress.com/bioconductor/paper2/

Appendix E: Representative RS Metadata Schemas

Citation File Format (CFF)

Citation File Format[477] (CFF) is a human- and machine-readable plain text file format (`CITATION.cff`) in YAML[478] that provides citation metadata for software. Code developers can include the CFF file in their repositories to let others know how to correctly cite their software. CFF is currently supported by GitHub[479], Zenodo[480], and Zotero[481], etc. The CFF file is a simple format by which code authors can record metadata that can then be translated into other formats or schemas.

CodeMeta initiative

The goals of the CodeMeta project is to create a concept vocabulary, i.e. a minimal metadata schema for academic SW and code, which can be used to standardize the exchange of SW metadata across platforms (e.g. repositories) and organisations. CodeMeta used the crosswalk[482] to generate a set of SW metadata concepts (CodeMeta-2.0[483] and CodeMeta-1.0[484]), arranged into JSON-LD and XML, respectively. CodeMeta is currently working to update schema.org so that all CodeMeta terms are included in it, and CodeMeta could become a community that supports research software within schema.org, rather than an independent schema.

DataCite Metadata Schema

The DataCite Metadata Schema is a list of core metadata properties chosen for the complete and consistent identification of a resource for citation and retrieval purposes, along with recommended use instructions. The latest version is Metadata Schema 4.3[485,486] released in 2019.

---

477 https://github.com/citation-file-format/citation-file-format

478 http://www.yaml.org/spec/1.2/spec.html

479 https://docs.github.com/en/github/creating-cloning-and-archiving-repositories/creating-a-repository-on-github/about-citation-files

480 https://twitter.com/ZENODO_ORG/status/1420357001490706442

481 https://twitter.com/zotero/status/1420515377390530560

482 https://github.com/codemeta/codemeta/blob/master/crosswalk.csv

483 https://doi.org/10.5063/schema/codemeta-2.0

484 https://doi.org/10.5063/schema/codemeta-1.0

485 https://schema.datacite.org/meta/kernel-4.3/metadata.xsd

486 https://doi.org/10.14454/f2wp-s162

DOE CODE initiative

The Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) worked on a project called the DOE Code Metadata Model[487] that defined metadata used for SW.

OpenAIRE guidelines for SW metadata

The OpenAIRE Project created SW guidelines to define and implement local SW management policies in exposing SW metadata and making SW citable. OpenAIRE created crosswalks and mappings[488] from other initiatives to make the OpenAIRE repositories compliant with all others in terms of discovery and citation of SW.

Appendix F: Best Practices Related to RS

The use of the term 'best practices' has different interpretations in different communities. A recent article entitled 'Good Enough Practices in Scientific Computing" presents a set of good computing practices for every research individual to adopt, regardless of their level of computational skill.[489]

F1: RS Management Best Practices

The core principles for RS management are the FAIR principles (FAIR4RS) that were derived from the FAIR data principles and are being adapted and adopted for research objects other than research data. In FAIR4RS, considerations include how the original principles could be adapted for the RS case, how to promote and measure the adoption of FARI4RS, and how to support the implementation.

Similar to DMP, a Software Management Plan[490] (SMP) is highly recommended by some funding agencies[491,492], which defines and formalizes a set of structures and goals to understand the RS in terms of roles and responsibilities throughout the RS lifecycle from the development phase to

---

[487] https://github.com/doecode/software-metadata/blob/master/XMLSchema/MetadataSchema.xsd

[488] https://docs.google.com/spreadsheets/d/1mKs-Pg_JuLcpqEkQqlSCs2gGC7nEEbhxdTbIoGcU6NI/edit#gid=0

[489] https://doi.org/10.1371/journal.pcbi.1005510

[490] The SMP is also referred to as other terminologies, e.g. the Software Development Plan, SW Project Management Plan (SPMP), the Statement of Work (SOW) by CANARIE and the Department of Energy (DOE) in the US, etc.

[491] https://webarchive.nationalarchives.gov.uk/ukgwa/20201001001808/https://epsrc.ukri.org/files/funding/calls/2012/computational-science-and-engineering-software-for-the-future/

[492] https://www.nsf.gov/pubs/2016/nsf16532/nsf16532.htm

sustainability consideration. Some guidelines and templates[493,494,495] have been developed to provide basic elements of a SMP, ensuring RS is accessible and reusable in the short, medium, and long term.

During the active management of RS while developing and/or modifying RS, best practices[496,497,498] include use of scripted analyses (e.g. R and Python) over mouse-operated point-and-click interface with commercial software (e.g. Microsoft's Excel, IBM's SPSS), following style guidelines (e.g. PEP8[499] for Python), code commenting and documentation (e.g. README file, automated documentation tools such as Sphinx[500] and Doxygen[501]), testing (e.g. `pytest`[502]), versioned releases (e.g. git), continuous integration (e.g. Travis CI[503]), including metadata files in the source code (e.g CodeMeta, CFF) and adding a registered license (e.g. SPDX, REUSE project).

Nowadays, researchers and research software developers rely on existing RS components and infrastructures for RS management. These RS infrastructures include IDEs (e.g. Jupyter notebook), workflow management systems (e.g. kepler, pegasus), archives (e.g.Software Heritage[504], Zenodo, institutional repositories), catalogs/registries (e.g, swMath, ASCL[505], OpenAire, ScanR, CRAN, PyPi, RunMyCode[506]), and publishers (e.g. IPOL, eLife, JOSS), which are also good solutions and practices[507] to improve the FAIRness of software, as well as to communicate and support software's dissemination, providing better possibilities for RS curation and sharing.

---

[493] 10.5281/zenodo.1422656

[494] https://cs.uwaterloo.ca/~apidduck/se362/Assignments/A2/spmp.pdf

[495] https://sceweb.uhcl.edu/helm/RationalUnifiedProcess/webtmpl/templates/mgmnt/rup_sdpln_sp.htm

[496] Best Practices for Scientific Computing: https://doi.org/10.1371/journal.pbio.1001745

[497] Good enough practices in scientific computing: https://doi.org/10.1371/journal.pcbi.1005510

[498] Software carpentry: https://software-carpentry.org/

[499] https://www.python.org/dev/peps/pep-0008/

[500] Sphinx-doc.org

[501] https://www.doxygen.nl/index.html

[502] https://docs.pytest.org/en/6.2.x/

[503] https://travis-ci.org/

[504] https://www.softwareheritage.org/howto-archive-and-reference-your-code/

[505] Astrophysics Source Code Library (ASCL): https://ascl.net/

[506] http://www.runmycode.org/

[507] During the years 2019-2020, the Task Force on Best Practices for Software Registries of the FORCE11 Software Citation Implementation Working Group worked to create Nine Best Practices for Scientific Software Registries and Repositories.

F2: Computational Reproducibility Best Practices (as Researchers or Data/Code authors)

Computational reproducibility, more specifically, defines computational reproducibility as "obtaining consistent results using the same input data, computational methods, and conditions of analysis"[508]. A new guide has been published attempting to answer the question "How reproducible should research software be?"[509], which defines four levels of reproducibility, suggesting criteria to help decide which level your RS should be at, and recommending practices to reach these levels of reproducibility.

A high-level principle of computational reproducibility is to provide a clear, specific, and complete description of how a reported result was reached, although different areas of study or types of inquiry may require different kinds of information. A computationally reproducible research package may include all or some of the following:

- Primary data (and documentation) collected and used in analysis
- Secondary data (and documentation) collected and used in analysis
- Primary data output result(s) (and documentation) produced by analysis
- Secondary data output result(s) (and documentation) produced by analysis
- Software program(s) (and documentation) for computing published results
- Research software program(s) for reproducing published results
    - Document RS in operation and make that documentation available
        - Version control
        - Capture running/computational environment:
            - including features of hardware (e.g. the numbers of cores in any CPUs) and features of software (e.g. the operating system, programming languages, supporting packages, other pieces of installed software, along with their versions and configurations)
        - Document SW dependencies, including
            - Installed SW, Integrated libraries and their associated versions
            - Metadata and specifications
            - The structure of source code

---

[508] National Academies of Sciences, Engineering, and Medicine. (2019). Reproducibility and replicability in science. National Academies Press.

[509] https://zenodo.org/record/4761867#.YKfD4ZNKg-Q

  - Individual components that support functionality
  - Parameters
  - Build and execution environment
- RS documentation and implementation details with high quality
- RDM and ensure full access to the data used in research
- Record computational research workflow and provenance information
- Published journal article

There are many resources available that aim to equip researchers with tools, knowledge, training, and confidence to follow reproducible open science good practice. Some examples include:

- The Turing Way
- R open science reproducibility guide
- (Cloud-based) reproducibility tools: Docker container, Binder, Code Ocean, Whole Tale, eLife Reproducible Document Stack, Galaxy, Gigantum, Manuscripts, o2r, REANA, ReproZip, Sweave, Stencila, Renku, binder, EaaSI, ReScience, Jupyter notebook, etc.

In addition to platforms and resources, social practices like grassroots communities (e.g. UK Reproducibility Network[510]) have emerged.

Appendix G: RS Management Through the Lens of COVID-19

The novel coronavirus disease (COVID-19) was officially declared as a global pandemic by the World Health Organization (WTO) on March 11, 2020[511]. At the time of writing this report, massive resources were dedicated to allowing researchers around the globe to tackle this crisis from multiple angles, including medical research, societal norms, community participation, data, computing power, and software, which resulted in extraordinary research outcomes through greater adoption of open science practices. Taking full advantage of this landscape, the response of both national and international research communities to help resolve this crisis provides a prime example for the impact of strong RS management practices and support. Examples include:

---

510 https://www.bristol.ac.uk/psychology/research/ukrn/

511 https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

- International funding agencies announced rapid research response to fund COVID-19 related research, most with requirements for related research findings (data, code, and publications) to be shared openly[512,513,514]. In Canada, this includes rapid response funding programs from the Tri-Agencies[515, 516].

- Global individual institutions and corporations also increased their investment in advanced research computing due to COVID-19. For example, in order to maximize the impact of the research, AMD initiated “a working group for COVID-19 HPC Fund recipients and AMD engineers to jointly discuss research areas and findings as well as hardware and software optimisations that can accelerate their collective work.”[517] Google also announced Google Cloud COVID-19 Research Grants.[518, 519] The COVID-19 Performance Computing (HPC) Consortium is another example that brought together the federal government, industry, and academic leaders to provide access to the world’s most powerful high-performance computing resources in support of COVID-19 research.[520]

- Research communities are collaborating to develop best practices for the sharing and reuse of COVID-19 research outcomes. Internationally, the Research Data Alliance (RDA) COVID-19 Working Group gathered global multidisciplinary experts to develop guidelines and recommendations on data sharing under the present COVID-19 circumstances.[521] The RDA COVID-19 Recommendations and Guidelines included a section highlighting best practices for research software. These recommendations focus on sharing and the key role it plays in data analysis and advancing research.

- Research sharing platforms (such as repositories) and curation services have joined forces to support research related to COVID-19. One of the most prominent is the Zenodo COVID-19 Community data repository, which is free and open for researchers worldwide to share a wide range of research outputs (publications, data, software, protocols, etc.) related to the COVID-19 epidemic or the SARS-CoV-2 virus.[522] Curation of these datasets is being supported by Europe’s OpenAIRE program, who has launched an online beta version of

---

[512] https://wellcome.org/coronavirus-covid-19/open-data

[513] https://coronavirus.frontiersin.org/covid-19-research-funding-monitor

[514] https://nebigdatahub.org/covid19/

[515] https://www.idrc.ca/en/news/funding-opportunity-rapid-research-response-novel-coronavirus-outbreak

[516] https://cihr-irsc.gc.ca/e/51890.html

[517] https://www.amd.com/en/press-releases/2020-09-14-amd-covid-19-hpc-fund-adds-18-institutions-and-five-petaflops

[518] https://edu.google.com/programs/credits/research/?modal_active=none

[519] https://uwaterloo.ca/research/google-apply-google-cloud-research-credits

[520] https://covid19-hpc-consortium.org/

[521] RDA COVID-19 Recommendations and Guidelines for Data Sharing

[522] https://zenodo.org/communities/covid-19/?page=1&size=20

COVID-19 Open Research Gateway[523] for aggregating, linking and providing access to COVID-19 scientific products across multiple scholarly communication infrastructures and sources. Research Data Canada (RDC) intended to gather software tools/platforms with a focus on COVID-19 in their Canadian COVID-19 Data Sharing Network Mindmap.[524]

- Research communities have also responded to the crisis. Examples of Canada's contribution to this effort, focusing on both human and virus genomics around COVID-19/SARS-CoV-2, are the iReceptor platform[525] (based at Simon Fraser University) and CanCoGen's VirusSeq Portal[526] (part of Genome Canada's CanCOGeN COVID-19 research initiative). iReceptor provides a Science Gateway[527] for exploring The AIRR Data Commons, a distributed network of standards-based data repositories that store sequencing data from the human antibody/B-cell and T-cell immune response (AIRR-seq data). Using a framework that follows the fundamental tenets of FAIR software development, the iReceptor team was able to extend the platform to curate, store, and share AIRR-seq data from a range of international COVID-19 studies.[528] As of May 28, 2021 over 1 billion annotated sequences from 15 COVID-19 studies are available in the AIRR Data Commons[529]. CanCoGen's VirusSeq Portal provides genomic-based tracking and analysis of the evolving traits of the SARS-CoV-2 virus across Canada, with the goals of identifying and tracking transmission trends at the regional, provincial, national and international scales, aiding detection of new clusters of cases/outbreaks, and discovering evolving viral characteristics. As of April 27, 2021, the VirusSeq Portal contains 111 viral genomes from two studies in Canada.

COVID-19 has demonstrated what "can" happen in the research world when there is a driving need. By looking at the progress that has been made in terms of scientific research into COVID-19, the publication of preprints with a very iterative open science approach has played an important role. The funding agencies have been able to act in a very agile way (prioritized resources for COVID work in *weeks* as opposed to the usual *months*) to support important and urgent COVID-19 research. This has been possible because of pre-existing partnerships between

---

[523] https://beta.covid-19.openaire.eu/

[524] https://share.mindmanager.com/#publish/Zv6lhDk-aGEptiZFkhjDBPrt7HTzvX6vvj8bEpcx

[525] Corrie et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories, Immunological Review, Volume 284, Issue1. https://doi.org/10.1111/imr.12666

[526] https://virusseq-dataportal.ca/ Canadian VirusSeq Data Portal (CanCOGeN VirusSeq) provides a vital link between Canada's public health units and researchers tracking the evolution of the virus and variants of concern.

[527] https://gateway.ireceptor.org/login

[528] Scott et al. The adaptive immune receptor repertoire community as a model for FAIR stewardship of big immunology data, Current Opinion in Systems Biology, Volume 24, December 2020 https://doi.org/10.1016/j.coisb.2020.10.001

[529] http://www.ireceptor.org/covid19

the funding agencies and the research communities. Another key factor has been the ability to use digital technologies remotely.

In addition, it has been compelling to see the cohesive strength of all stakeholders (including research communities, funders) and the speed at which they were able to come together; to get things off the ground; to get access to computing facilities; and the speed at which they have been able to deploy RS and RS platforms.

Although it is too soon to evaluate the full impact of those initiatives and related programs on the pandemic itself, it marks a clear mindset shift in the DRI ecosystem towards greater openness and collaboration, as well as adoption of best practices in the open science movement of national and international importance. In terms of the RS landscape in Canada, without Canada's continued investment in best practices of RS development, and a commitment to ensuring Canada is a leader in the RS community, these developments would not happen. More importantly, the challenge is how can the level of collaboration and data sharing that we have seen around COVID-19 be incentivized so that it continues post pandemic and can we take advantage of this opportunity to help coordinate and advance this collaboration.

Appendix H: Funders for RS Development and Use

Name	Website	Region/type	Notes
Agriculture and Agri-Food Canada: AgriRisk Initiatives: Micrograants	http://www.agr.gc.ca/eng/agricultural-programs-and-services/agririsk-initiatives-micrograants/?id=1531512971010	Federal/Government funding sources	
Alberta Cancer Foundation	https://www.albertacancer.ca/	Provincial/regional governments/development agencies (AB)	
Alberta Centre for Child, Family and Community Research (ACCFCR)	https://policywise.com/	Provincial/regional governments/development agencies (AB)	
Alberta Innovates (AB Innov)	https://albertainnovates.ca/	Provincial/regional governments/development agencies (AB)	
Alzheimer Society of Canada (ASC)	https://alzheimer.ca/en	Foundations & corporations (public)	
Amazon Research Awards	https://www.amazon.science/research-awards	Private company investment	
Andrew W. Mellon Foundation	https://mellon.org/	Foundations & corporations (private)	One of the US foundation grants tjat have funded some HSS (humanities and social sciences) and RS

			development
Arthritis Society	https://arthritis.ca/	Foundations & corporations (public)	
Atlantic Canada Opportunities Agency (ACOA)	https://www.canada.ca/en/atlantic-canada-opportunities.html	Provincial/regional governments/development agencies (Atlantic)	
Banting Research Foundation	https://www.bantingresearchfoundation.ca/	Foundations & corporations (public)	
Bill and Melinda Gates Foundation (BMGF)	https://www.gatesfoundation.org/	International funding (private)	Based in the US
Brain Canada	http://www.caro-acro.ca/	Foundations & corporations (public)	
Brain Tumour Foundation of Canada	https://www.braintumour.ca/	Foundations & corporations (public)	
C17 Council (C17)	http://www.c17.ca/	Foundations & corporations (public)	
Canada Excellence Research Chairs (CERC)	https://www.cerc.gc.ca/home-accueil-eng.aspx	Federal/Government funding for strategic initiatives	
Canada First Research Excellence Fund (CFREF)	https://www.cfref-apogee.gc.ca/home-accueil-eng.aspx#	Federal/Government funding for strategic initiatives	e.g., Global Water Futures - University of Saskatchewan
Canada Research Chairs Program (CRCP)	https://www.cerc.gc.ca/home-accueil-eng.aspx	Federal/Government funding for strategic initiatives	
Canada Research Continuity Emergency Fund (CRCEF)	https://www.sshrc-crsh.gc.ca/funding-financement/crcef-fucrc/index-eng.aspx	Federal/Government funding for strategic initiatives	
Canada's National Research and Education Network (NREN)	https://www.canarie.ca/network/nren/	Consortia	
Canadian Association of Research Libraries (CARL)	https://www.carl-abrc.ca/	Consortia	
Canadian Blood Services	https://www.blood.ca/en	Foundations & corporations (public)	
Canadian Cancer Society (CCS)	https://www.cancer.ca/	Foundations & corporations (public)	
Canadian Foundation for Innovation (CFI)	https://www.innovation.ca/	Federal/Government funding for strategic initiatives	
Canadian Hemophilia Society (CHS)	https://www.hemophilia.ca/	Foundations & corporations (public)	
Canadian Institutes of Health Research (CIHR)	http://www.cihr-irsc.gc.ca/e/193.html	Federal/Government funding sources	
Canadian International Innovation Program (CIIP)	https://www.tradecommissioner.gc.ca/funding-financement/ciip-pcii/index.aspx?lang=eng	Federal/Government funding for strategic initiatives	
Canadian Northern Economic Development Agency (CanNor)	https://www.cannor.gc.ca/eng/1351104567432/1351104589057	Federal/Government funding sources/Indigenous communities/agencies	

Canadian Orthopaedic Foundation	https://whenithurtstomove.org/	Foundations & corporations (public)	
Canadian Partnership Against Cancer (CPAC)	https://www.partnershipagainstcancer.ca/	Foundations & corporations (public)	
Canadian Prostate Cancer Research Initiative (CPCRI)	https://webapps.cihr-irsc.gc.ca/cris/agency_info_E?p_agency_cd=CPCRI&p_popup=T	Foundations & corporations (public)	
Canadian Space Agency (CSA)	https://www.asc-csa.gc.ca/eng/Default.asp	Federal/Government funding sources	
Canadian Tobacco Control Research Initiative (CTCRI)	http://www.ohpe.ca/node/4717	Foundations & corporations (public)	
CANARIE	https://canarie.ca/	Federal/Government funding for strategic initiatives	
Cancer Care Nova Scotia (CCNS)	http://www.nshealth.ca/cancer-care	Provincial/regional governments/development agencies (NS)	
Cancer Care Ontario (CCO)	https://www.cancercareontario.ca/en	Provincial/regional governments/development agencies (ON)	
Cancer Research Society (SRC)	https://www.societederecherchesurlecancer.ca/en	Foundations & corporations (public)	
CancerCare Manitoba (CCMB)	https://www.cancercare.mb.ca/Patient-Family/cancercare-team	Provincial/regional governments/development agencies (MB)	
Children's Oncology Group (COG)	https://childrensoncologygroup.org/	Foundations & corporations (public)	The Children's Oncology Group is primarily funded by the National Cancer Institute and also receives additional funding from other granting agencies and from philanthropic sources (thecogfoundation.org)
Colleges and Institutes Canada (CICan)	https://www.collegesinstitutes.ca/	Federal/Government funding sources/Indigenous communities/agencies	
Colleges and Institutes Canada (CICan) - Clean Tech Internships	https://careerlauncher.ca/cleantech/?utm_source=email&utm_medium=referral&utm_campaign=CICanOutreach	Federal/Government funding sources/Indigenous communities/agencies	
Communitech DATA.BASE program	https://www.communitech.ca/	Foundations & corporations (public + private)	
Consortium de recherche et d'innovation en aérospatiale au Québec (CRIAQ)	https://www.criaq.aero/	Consortia	
Consortium de RECHERCHE et d'INNOVATION en transformation métallique (CRITM)	https://www.critm.ca/	Consortia	
Consortium de recherche et	http://www.cribiq.qc.ca/en/	Consortia	

innovations en bioprocédés industriels au Québec (CRIBIQ)			
Consortium Québécois sur la Découverte du Médicaments (CQDM)	https://cqdm.org/en/	Consortia	
Council on Library and Information Resources (CLIR)	https://www.clir.org/	Foundations & corporations (public)	Based in the US, the CLIR Postdoctoral Fellowship Program funds software curation (e.g., for virtual reality and 3D imaging applications)
Crohn's and Colitis Canada	https://crohnsandcolitis.ca/	Foundations & corporations (public)	
Crown-Indigenous Relations and Northern Affairs Canada-Clean Tech Internships	https://www.canada.ca/en/crown-indigenous-relations-northern-affairs.html	Federal/Government funding sources/Indigenous communities/agencies	
Crown-Indigenous Relations and Northern Affairs Canada-Clean Tech Internships - Northern Contaminants Program (NCP)	http://science.gc.ca/eic/site/063.nsf/eng/h_98204.html	Federal/Government funding sources/Indigenous communities/agencies	
Department of Canadian Heritage	https://www.canada.ca/en/canadian-heritage.html	Federal/Government funding for strategic initiatives	
Department of Canadian Heritage - Canada History Fund	https://www.canada.ca/en/canadian-heritage/services/funding/canada-history-fund/program-details.html	Federal/Government funding for strategic initiatives	
Department of National Defence (DND)	https://www.canada.ca/en/department-national-defence.html	Federal/Government funding for strategic initiatives	
Dystonia Medical Research Foundation Canada (DMRFC)	https://dystoniacanada.org/	Foundations & corporations (private)	
Early Researcher Awards	https://www.ontario.ca/page/early-researcher-awards	Provincial/regional governments/development agencies (ON)	
Employment and Social Development Canada	https://www.canada.ca/en/employment-social-development.html	Federal/Government funding for strategic initiatives	
Environment and Climate Change Canada (ECCC)	https://www.canada.ca/en/environment-climate-change.html	Federal/Government funding for strategic initiatives	
European Commission "Horizon 2020"	https://ec.europa.eu/programmes/horizon2020/en	International funding (public)	
Federation of Canadian Municipalities (FCM), Green Municipal Fund	http://www.fcm.ca/	Foundations & corporations (private)	
FRQ - Nature et Technologies (FRQNT)	http://www.frqnt.gouv.qc.ca/en/	Provincial/regional governments/development agencies (QC)	
FRQ - Santé (FRQS)	http://www.frqs.gouv.qc.ca/en/	Provincial/regional governments/development	

		agencies (QC)	
FRQ - Société et Culture (FRQSC)	http://www.frqsc.gouv.qc.ca/en/	Provincial/regional governments/development agencies (QC)	
Genome BC	https://www.genomebc.ca/	Provincial/regional governments/development agencies (BC)	
Genome Canada	https://www.genomecanada.ca/en/	Federal/Government funding for strategic initiatives	
Génome Québec	https://www.genomequebec.com/	Provincial/regional governments/development agencies (QC)	
Global Affairs Canada (GAC)	https://www.international.gc.ca/global-affairs-affaires-mondiales/home-accueil.aspx?lang=eng	International funding (public)	
Google Cloud research credits	https://edu.google.com/intl/en_ca/programs/credits/research/?modal_active=none	Private company investment	
Gordon and Betty Moore Foundation	https://www.moore.org/	International funding (private)	Based in the US
Gordon Foundation	https://gordonfoundation.ca/	Foundations & corporations (private)	
Government of Canada	https://www.canada.ca/en/employment-social-development/services/funding.html	Federal/Government funding sources	
Graham Boeckh Foundation (GBF)	https://grahamboeckhfoundation.org/en/	Foundations & corporations (private)	
Grand Challenges	https://grandchallenges.org/grant-opportunities	International funding (public)	
Grand Challenges Canada (GCC)	https://grandchallenges.org/grant-opportunities	International funding (public)	
Grants.gov	Grants.gov	International funding (public)	US grant programs in the US. e.g., Department of Defense (DOD), National Institutes of Health (NIH), National Science Foundation (NSF), Department of Energy (DOE)
Health Canada	https://www.canada.ca/en/health-canada.html	Federal/Government funding for strategic initiatives	
Heart And Stroke Foundation	https://www.heartandstroke.com/	Foundations & corporations (public)	
Helmsley Charitable Trusthelmsleytrust.org	https://helmsleytrust.org/	International funding (private)	Based in the US
Huawei Innovation Research Program (HIRP)	https://innovationresearch.huawei.com/IPD/hirp/portal/index.html	Private company investment	
Impact Canada	http://impact.canada.ca/en/node/19	Federal/Government funding for strategic initiatives	
Innovation Super Clusters	https://www.ic.gc.ca/eic/site/093.nsf/eng/home	Federal/Government funding for strategic initiatives	

Innovation, Science and Economic Development Canad (ISED)	http://strategis.ic.gc.ca/home	Federal/Government funding for strategic initiatives	
InnovÉÉ	https://innov-ee.ca/	Consortia	
International Development Research Centre (IDRC)		International funding (public)	
Jarislowsky Foundation	https://www.charitydata.ca/charity/the-jarislowsky-foundationla-fondation-jarislowsky/894744036RR0001/	Foundations & corporations (public)	
McLean Foundation	http://www.mcleanfoundation.ca/	Foundations & corporations (private)	
Mental Health Research Canada (MHRC)	https://www.mhrc.ca/	Foundations & corporations (public)	
Michael J. Fox Foundation (MJFF)	https://www.michaeljfox.org/	International funding (public)	Based in the US
Michael Smith Foundation for Health Research (MSFHR)	https://www.msfhr.org/	Foundations & corporations (public); Provincial/regional governments/development agencies (BC)	
Ministère de l'Économie et de l'Innovation (MEI) (formerly MESI)	https://www.economie.gouv.qc.ca/	Provincial/regional governments/development agencies (QC)	
Ministère des relations internationales et de la Francophonie (MRIF)	http://www.mrif.gouv.qc.ca/en/	Provincial/regional governments/development agencies (QC)	
Ministry of Colleges and Universities	https://www.ontario.ca/page/ministry-colleges-universities	Provincial/regional governments/development agencies (ON)	
Ministry of Research, Innovation and Science (MRIS)	https://www.ontario.ca/page/ministry-economic-development-job-creation-trade	Provincial/regional governments/development agencies (ON)	
MITACS	https://www.mcgill.ca/research/research/funding/Federal/Government/mitacs	Federal/Government funding for strategic initiatives	
National Research Council Canada (NRC)	https://nrc.canada.ca/en	Federal/Government funding for strategic initiatives	
Natural Resources Canada (NRCan)	https://www.nrcan.gc.ca/home	Federal/Government funding for strategic initiatives	
Natural Sciences and Engineering Research Council of Canada (NSERC)	http://www.nserc-crsng.gc.ca/index_eng.asp	Federal/Government funding sources	
Networks of Centres of Excellence (NCE)	https://www.nce-rce.gc.ca/Index_eng.asp	Federal/Government funding for strategic initiatives	
New Brunswick Health Research Foundation (NBHRF)	https://www.nbhrf.com/	Provincial/regional governments/development agencies (NB)	
New Frontiers in Research Fund	https://www.sshrc-crsh.gc.ca/funding-	Federal/Government funding sources	

(NFRF)	financement/nfrf-fnfr/index-eng.aspx		
Nova Scotia Health Research Foundation (NSFRF)	http://www.cdha.nshealth.ca/discovery-innovation/research-fund	Provincial/regional governments/development agencies (NS)	
Ocean Protection Funding | Waitt Foundation	https://www.waittfoundation.org/	Foundations & corporations (private)	Headquartered in Washington, DC
Ontario Brain Institute (OBI)	https://braininstitute.ca/	Provincial/regional governments/development agencies (ON)	
Ontario Centres of Excellence (OCE)	https://www.oce-ontario.org/	Provincial/regional governments/development agencies (ON)	
Ontario COVID-19 Rapid Research Fund	https://www.ontario.ca/page/ontario-covid-19-rapid-research-fund	Provincial/regional governments/development agencies (ON)	
Ontario Genomics	https://www.ontariogenomics.ca/	Provincial/regional governments/development agencies (ON)	
Ontario Institute for Cancer Research (OICR)	https://oicr.on.ca/	Provincial/regional governments/development agencies (ON)	
Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA)	http://www.omafra.gov.on.ca/english/	Provincial/regional governments/development agencies (ON)	
Ontario Ministry of Health	https://www.ontario.ca/page/ministry-health	Provincial/regional governments/development agencies (ON)	
Ontario Ministry of Transportation (MTO)	https://www.ontario.ca/page/ministry-transportation	Provincial/regional governments/development agencies (ON)	
Ontario Research Fund – Research Excellence (ORF-RE)	https://www.ontario.ca/page/ontario-research-fund-research-excellence	Provincial/regional governments/development agencies (ON)	
Ontario Research Fund – Research Infrastructure	https://www.ontario.ca/page/ontario-research-fund-research-infrastructure	Provincial/regional governments/development agencies (ON)	
Ovarian Cancer Canada (NOCA)	https://ovariancanada.org/	Foundations & corporations (public)	
Pancreatic Cancer Canada Foundation (PCCF)	https://pancreaticcancercanada.ca/	Foundations & corporations (public)	
Paper Presentation Grants	https://www.mcgill.ca/research/research/funding/internal	Internal Funding	
Partnership for Clean Competition (PCC)	https://cleancompetition.org/	Foundations & corporations (public)	Based in the US
Polar Knowledge Canada	https://www.canada.ca/en/polar-knowledge.html	Federal/Government funding sources/Indigenous communities/agencies	
PRIMA Québec (formerly NanoQuébec)	https://www.prima.ca/en	Consortia	

PROCURE	https://www.procure.ca/en/home/	Foundations & corporations (private)	
PROMPT (formerly CINQ)	https://promptinnov.com/en/	Consortia	
Prostate Cancer Canada (PCC)	https://www.prostatecancer.ca/	Foundations & corporations (public)	
Public Health Agency of Canada (PHAC)	https://www.canada.ca/en/public-health.html	Federal/Government funding sources	
Public Safety Canada	https://www.publicsafety.gc.ca/index-en.aspx	Federal/Government funding for strategic initiatives	
Quebec Breast Cancer Foundation (QBCF)	https://www.charityintelligence.ca/charity-details/612-quebec-breast-cancer-foundation	Foundations & corporations (public)	
Research ManitobaResearch Manitoba (MHRC)	https://researchmanitoba.ca/	Provincial/regional governments/development agencies (MB)	
Royal Society of Canada (RSC)	https://rsc-src.ca/en	Foundations & corporations (public)	
Sabbatical Leave Research Grants	https://www.mcgill.ca/research/research/funding/internal	Internal Funding	
Saskatchewan Cancer Agency	http://www.saskcancer.ca/	Provincial/regional governments/development agencies (SK)	
Saskatchewan Health Research Foundation (SHRF)	https://www.shrf.ca/	Provincial/regional governments/development agencies (SK)	
Shastri Indo-Canadian Institute (SICI)	https://www.shastriinstitute.org/	Foundations & corporations (public)	Binational organization
Simons Foundation	https://www.simonsfoundation.org/	Foundations & corporations (private)	
Sloan Foundation	https://sloan.org/	International funding (private)	Based in the US
Social Sciences and Humanities Research Council of Canada (SSHRC)	http://www.sshrc-crsh.gc.ca/home-accueil-eng.aspx	Federal/Government funding sources	
SSHRC General Research Fund (GRF) and NSERC GRF	https://uwaterloo.ca/research/find-and-manage-funding/manage-funding/closing-research-accounts/general-research-fund-guidelines-nserc-and-sshrc-only	Internal Funding	
Status of Women Canada	https://cfc-swc.gc.ca/index-en.html	Federal/Government funding for strategic initiatives	
Technology Academy Finland (TAF)	https://taf.fi/	International funding (public)	
The Impact Assessment Agency of Canada	https://www.canada.ca/en/impact-assessment-agency.html	Federal/Government funding for strategic initiatives	
The Institute for Catastrophic Loss Reduction (ICLR)	https://www.iclr.org/	Foundations & corporations (public)	
The Terry Fox Foundation (TTF)	https://terryfox.org/	Foundations & corporations (public)	

Transport Canada	https://tc.canada.ca/en/node/22	Federal/Government funding for strategic initiatives	
U.S. Air Force - Research Interests of the Air Force Office of Scientific Research	https://www.Federal/Governmentgrants.com/Research-Interests-of-the-Air-Force-Office-of-Scientific-Research-68274.html	International funding (public)	US Federal/Government grants
UK Research and Innovation (UKRI)	https://www.ukri.org/	International funding (public)	Based in the UK
Wellcome Trust	https://wellcome.org/	International funding (private)	Based in the US
Weston Family Foundation	https://www.westonfoundation.org/	Foundations & corporations (private)	
Workplace Safety and Insurance Board Ontario (WSIB)	https://www.wsib.ca/en/grants-program	Provincial/regional governments/development agencies (ON)	
NumFOCUS	https://numfocus.org/	US	SW development

Appendix I: RS Community Landscape

Name	Type	Website	Region	Discipline	Software Focus	Notes
(International) Society of Research Software Engineering/UK-RSE	Community initiative, events	https://society-rse.org/	global	all	general	
ACM Badging schema	Projects	https://www.acm.org/publications/policies/artifact-review-and-badging-current	global	all	Artifact evaluation, policy	ACM. (2016). Artifact review and badging. https://www.acm.org/publications/policies/artifact-review-and-badging ⇒ Artifact Review and Badging Version 1.1 - August 24, 2020
Alliance of German Science Organisations: Working group on Software and Services		https://www.allianzinitiative.de/fields-of-action-projects/digital-tools-software-and-services/?lang=en	Germany	all	general	search software organisations, initiatives, programs, projects, grants, communities of practice, networks, journals, conferences

Allianz-Initiative	Government/community initiative	https://www.allianzinitiative.de/fields-of-action-projects/digital-tools-software-and-services/?lang=en	Germany	all	general	
Apache Software Foundation	Community initiative	https://www.apache.org/	global	all	open science	
Astrophysics Source Code Library	Registry	https://ascl.net/	global	astronomy	indexing, citation	
Australian Research Data Commons	Government initiative	https://ardc.edu.au/collaborations/strategic-activities/software-and-platforms/	Australia	all	software development and sustainability	
Awesome Research Software Registries	Registry, catalog, index, repository, platform	https://github.com/NLeSC/awesome-research-software-registries	global	all	registry,	
Better Scientific Software (BSSw)	Community initiative	http://bssw.io	USA, international	all	sustainability, productivity	
Canadian Research Software Conference (CRSC)	Events	https://researchsoftware.ca/	Canada	all	general, community building, sustainability	
CANARIE (Canada)	Government initiative	https://www.canarie.ca/language/	Canada	all	sustainability	
Code.gov	Platform, program, registry	https://code.gov/	global	all	good practices, open source, sharing, collaboration, reuse,	
CodeMeta initiative	Community initiative, groups	https://codemeta.github.io/	global	all	Metadata, citation	
Community Surface Dynamics Modeling System (CSDMS)	Community initiative	https://csdms.colorado.edu/	global	geosciences	general	

CRAN	registry	https://cran.r-project.org/	global	R code	package mangers	
Dagstuhl	Journal, projects	https://www.dagstuhl.de/en/publications/oasics	Germany	Informatics	publishing	
Dagstuhl Artifacts Series (DARTS)	Projects	https://www.dagstuhl.de/en/publications/darts/	Germany	all	Artifact evaluation, policy	Wagner, M. (2017). Hitting the Bull'S eye with darts: Artifact evaluation in computer science. https://doi.org/10.5281/ZENODO.583007
DARPA Open Catalog	Catalog	http://www.darpa.mil/opencatalog/	USA	all	curation, registry,	
de-RSE	RSE group	https://www.de-rse.org/en/	Germany	all	general	
de-RSE - Society for Research Software Engineers	Community initiative	https://de-rse.org/de/index.html	Germany	all	general	
Debian Astro	Community initiative	https://blends.debian.org/astro	global	astrophysics	productivity	
DEVLOG network	Community initiative, groups	http://devlog.cnrs.fr/region	France	all	community building, general	
Digital Preservation Coalition	Community initiative	https://www.dpconline.org/	global	all	sustainability	has some software events: https://www.dpconline.org/knowledge-base/preservation-lifecycle/software
EaaSI	Not-for-profit entity, Community initiative, Program, Project, Grant, Community of practice, Network	https://www.softwarepreservationnetwork.org/eaasi/	USA, international	all	preservation	
eLife	Not-for-profit entity, Community initiative, Journal, Project, Community of practice	https://elifesciences.org/	global	biomedical, biochemistry, life sciences	open source, sharing, culture, practices, collaboration, computational	

					reproducibility, EDI, productivity	
ELIXIR	International government initiative	https://elixir-europe.org/	global	Life sciences	registry of tools, training	
EOSC Architecture WG - Scholarly Infrastructures for Research Software (SIRS)	Community initiative, groups	https://www.eoscsecretariat.eu/working-groups/architecture-working-group	Europe	all	infrastructure, sustainability, interoperability, PID,	Towards an open architecture of scholarly infrastructures for research software
EOSC-Context: OpenAireAdvanced, EOSChub, EOSCpilot	International government initiative	https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud	Europe	all	interoperability, registry of tools	
ePLAN	Community initiative	https://escience-platform.nl/	Netherlands	all	general	
European Environment for Scientific Software Installations (EESSI)		https://www.eessi-hpc.org/	Europe			
F1000Research	Groups, publisher, platform	https://f1000research.com/	global	all	general	new group on "code review". F1000 code review effort Working Group plan
Five recommendations for FAIR software	Project	https://fair-software.nl/	global	domain agnostic	general	
FORCE11 Software Citation Implementation Working Group	Community initiative	https://www.force11.org/group/software-citation-implementation-working-group	global	all	citation	
German Research Foundation	Government initiative	https://www.dfg.de/en/index.jsp	Germany	all	sustainability	
Helmholtz Open Science	Not-for-profit entity	https://os.helmholtz.de/open-science-in-der-helmholtz-gemeinschaft/forschungssoftware/	Germany	natural sciences, life sciences, engineering	open science	
Hidden REF	Community initiative	https://hidden-ref.org/	UK	all	citation	

Hidden Ref	Community initiative	https://hidden-ref.org/	UK	all	impact	
HUBzero	Platform	https://hubzero.org/	global	all	website development, open source, publishing, sharing, scientific collaboration, community building, computing and analytical tools,	
IDEAS Software Productivity	Project	https://ideas-productivity.org/	USA	all	sustainability, productivity	
Image Processing On Line (IPOL)	Journal	https://www.ipol.im/	global	image processing, image analysis, mathematics,	reproducibility, open science,	
INCF/OCNS Infrastructure/Software/Tools SIG	Not-for-profit entity, Community initiative	https://neurostars.org/t/ocns-infrastructure-software-tools-sig-meet-and-greet-initial-discussions/15560/17, https://github.com/orgs/OCNS/teams/software-sig	global	neuroscience	general	
Journal of Open Research Software (JORS)	Journal	https://openresearchsoftware.metajnl.com/	global	all	citation, preservation	
Journal of Open Source Software (JOSS)	Journal	http://joss.theoj.org	global	all	citation, preservation	
Knowledge Exchange (cooperation of Jisc, SURF, CSC, DFG, DEFF, CNRS)	Collaboration, community initiative	http://www.knowledge-exchange.info/	Europe	all	open access, open scholarship, sharing,	
Knowledge Exchange (cooperation of Jisc,		http://www.knowledge-	global	all	sustainability	

SURF, CSC, DFG, DEFF, CNRS)		exchange.info/				
Mass Open Cloud	Project	https://massopen.cloud	USA	all	Reproducibility, sharing,	
Mentored Sprints	Community initiative, groups	https://mentored-sprints.netlify.app/	USA	all	best practices, training	
Molecular Science Software Institute	Project	https://molssi.org	global	molecular science		
Mozilla	Company	https://www.mozilla.org/en-US/	global	all	open science	
NASA Open Source Software	Catalog	https://code.nasa.gov/	USA	all	sharing, registry,	
Netherlands eScience Center	Government initiative	https://www.esciencecenter.nl/	Netherlands	all	general	
Network for Computational Modeling in Social and Ecological Sciences (CoMSES Net)	Community initiative, platform, repository,	https://www.comses.net/	global	Social and Ecological Sciences	good practices, citation, preservation, preproduciblity, reuse	
New Zealand eScience Infrastructure (NeSI)	Government initiative	https://www.nesi.org.nz/	New Zealand	all	general	
NIH Data Commons	Government initiative	https://commonfund.nih.gov/commons	USA	all	interoperability	
NL-RSE	RSE group	https://nl-rse.org/	Netherlands	all	general	
Nordic-RSE Group	Community initiative	https://nordic-rse.org/	Nordic countries	all	general	
NumFOCUS	Not-for-profit entity	https://numfocus.org/	global	all	open science	
NumFOCUS	Not-for-profit entities	https://numfocus.org/	global	all	best practices, community building	
Open Bioinformatics Foundation	Community initiative	https://www.open-bio.org/	global	biosciences	general	
Open Force Field Initiative	Community initiative	https://openforcefield.org/	global	biosciences, chemistry,	software development &	

				materials, physics	interoperability	
OpenAIRE Explore	Platform, infrastructure, aggregator,	https://explore.openaire.eu/	Europe	all	interoperability, registry of tools, open science, sharing, discoverability, transparency, good practices,	
OpenForum Europe (OFE)	Not-for-profit entities	https://openforumeurope.org/	Europe	all	interoperability, open source, policy	
PLAN-E	Community initiative	https://plan-europe.eu/	EU	all	sustainability	
PyPI	registry	https://pypi.org/	global	Python code	package mangers	
RDA CURE-FAIR WG	Community initiative, groups	https://www.rd-alliance.org/groups/cure-fair-wg	global	all	Curation, reproducibility, sustainability, FAIR	
RDA FAIR for Research Software (FAIR4RS) WG	Community initiative, groups	https://www.rd-alliance.org/groups/fair-research-software-fair4rs-wg	global	all	FAIR, general, sustainaibility	
RDA Software Source Code Identification (SCID) WG	Community initiative, groups	https://www.rd-alliance.org/groups/rdaforce11-software-source-code-identification-wg	global	all	Identification, general	Software Source Code Identification Use cases and identifier schemes for persistent software source code identification
RDA Software Source Code IG (SSC IG)	Community initiative	https://www.rd-alliance.org/groups/software-source-code-ig	global	all	general	
RDA-COVID19-Software	Community initiative, groups	https://www.rd-alliance.org/groups/rda-covid19-software	global	all	sharing, sustainability, bes practices	RDA COVID-19 Recommendations and Guidelines for Data Sharing

Red Hat Collaboratory	Community initiative, companies	https://www.bu.edu/rhcollab/	USA	all	community building, open source	
ReSA (Research Software Alliance)	Community initiative	https://www.researchsoft.org/	global	all	general	Research Software Alliance (ReSA) Taskforce
Research Data Alliance	Community initiative	https://www.rd-alliance.org/	global	all	general	
Research Software Portal @CANARIE	Registry	https://science.canarie.ca/researchsoftware/home/main.html	Canada	all	registry, indexing sustainability	
rOpenSci	Community initiative	https://ropensci.org	global	all	general	
RSE Assoc of Aust/NZ (RSE-aunz)	Community initiative	https://rse-aunz.github.io/	Australia/New Zeland	all	general	
RSE Assoc of Belgium (be-RSE)	Community initiative	https://www.be-rse.org/	Belgium	all	general	
ScanR	Aggregator	https://scanr.enseignementsup-recherche.gouv.fr/	France	all	aggregator, indexing, general	
SciCodes	Community initiative	https://scicodes.net/	global	all	registries, repositories, best practices, reproducibiltiy, preservation, discoverabiltiy, citation	
Science Code Manifesto	Community initiative, projects	http://sciencecodemanifesto.org/, https://github.com/codeisscience/manifesto	global	all	citation, curation, licensing, good practices	Barnes, N. (2010). Science code manifesto. http://sciencecodemanifesto.org
Science Gateways Catalog	Registry, catalog	https://catalog.sciencegateways.org/	USA	all	science gateways, general, reuse,	
Science Gateways Community Institute (SGCI)	Community initiative / project	http://sciencegateways.org	USA	all (science gateways)	general	

Scientific Software Registry Collaboration Workshop	Community initiative	https://asclnet.github.io/SWRegistryWorkshop/Products/Products.html	global	all	registries, repositories, metadata standards, citation	
SE4Science	Community initiative	https://se4science.org/workshops/	global	all	open science, sustainability	
Software Heritage	Platform	https://www.softwareheritage.org	global	all	preservation, citation	
Software Impacts	Journal	https://www.sciencedirect.com/journal/software-impacts	global	all	citation, preservation	
Software Preservation Network (SPN)	Community initiative	https://www.softwarepreservationnetwork.org/	global	all	preservation	
Software Sustainability Institute (SSI)	Community initiative / project	https://www.software.ac.uk/	UK	all	general	
SoftwareX	Journal	https://www.sciencedirect.com/journal/softwarex	global	all	peer-review, reproducibility, citation	open access
SORSE (Series of Online Research Software Events)	Events	https://sorse.github.io	global	all	community building, collaboration,	
swMATH	Registry, Community initiative,	https://swmath.org/	global	mathematics	registry, open access, discoverability,documentation, related publication	
The Carpentries	Not-for-profit entity	https://carpentries.org/	global	all	software development	

The EinsteinPy Project	Project	https://einsteinpy.org	India, Europe	Gravitational Physics, General Relativity	sustainability, productivity, preservation	
UKRI (UK research councils) e-infrastructure expert group, RSE fellowships etc	Government initiative	https://www.ukri.org/	UK	all	sustainability, productivity	e-infrastructure expert group, RSE fellowships etc
URSSI	Government/community initiative	http://urssi.us/	USA	all	sustainability	
US-RSE Association	Community initiative	https://us-rse.org/	USA	all	general	
WSSSPE (Working towards Sustainable Software for Science: Practice and Experiences)	Community initiative, events	http://wssspe.researchcomputing.org.uk	global	all	general	
NeuroImaging Tools & Resources Collaboratory (NITRC)	Community, platform, initiative, inftrastructure, registry	https://www.nitrc.org/	USA	neuroinformatics	community building, collaobration, discovery, open access, good practices, repositories, package mangers, reproducibility	
Whole Tale	Project, platform, infrastructure	https://wholetale.org/	USA	all	open source, reproducibility, collaboration, discovery, publishing, sharing, reuse	